2023-08-22

情報幾何 –備忘録メモ–

関原 謙介
株式会社シグナルアナリシス

まえがき

これまでの情報幾何に関連したノートをまとめたものである．

各ノートには，いちいち別のノートを参照する手間を省くため，内容に重複が多い．このようにまとめるに際しても，重複はそのままにした．

各ノートは，見つかった間違いなどは修正した「最新版」である．

目 次

第1章　情報幾何–スカラー正規分布多様体–

オリジナル 2023-5-30, 間違い修正版 2023-5-31

情報幾何に対する理解を深めるため，簡単で，なじみ深いスカラー（1 次元）正規分布の多様体を考え，測地線，2 点間の最短距離，曲率などを（微分幾何学でふつう使われる）リーマン接続を用いて計算してみた．これらの計算は，情報幾何の特徴をなす考え方–双対接続，アルファ接続，双対平坦性–などを用いていないため，情報幾何の本流の議論ではない．

1.1　フィシャー計量

1.1.1　統計モデル

$p(\boldsymbol{x}|\boldsymbol{\theta})$ を確率変数 $\boldsymbol{x}$ でパラメータ $\boldsymbol{\theta} = (\theta^1, \dots, \theta^n)$ を持つ確率（密度）分布とする．確率密度分布は以下の条件を満たす.

$$\{p(\boldsymbol{x}|\boldsymbol{\theta}) : \Omega \to \mathbb{R} \,|\, p(\boldsymbol{x}|\boldsymbol{\theta}) > 0, \int_\Omega p(\boldsymbol{x}|\boldsymbol{\theta}) dx = 1\}$$

すると，

$$M = \{p_{\boldsymbol{\theta}} = p(\boldsymbol{x}|\boldsymbol{\theta}) \,|\, \boldsymbol{\theta} = (\theta^1, \dots, \theta^n) \in \Theta\}$$

を n 次元統計モデルとよぶ．この M は $\boldsymbol{\theta} = (\theta^1, \dots, \theta^n)$ を座標系とする多様体とみなす.「統計モデル M を $\boldsymbol{\theta} = (\theta^1, \dots, \theta^n)$ を座標系とする多様体と同一視する.」と言うこともできる.

例 (1)： スカラー正規分布 $\Omega = \mathbb{R}, n = 2$, $\boldsymbol{\theta} = (\mu, \sigma)$ として

$$p(x|\boldsymbol{\theta}) = p(x|(\mu, \sigma)) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

例 (2)： 任意の離散分布，$\Omega = \{x_1, \dots, x_n, x_{n+1}\}$, $\boldsymbol{\theta} = \{(\theta^1, \dots, \theta^n) \,|\, \theta^i > 0, \sum_{i=1}^n \theta^i < 1\}$ として，

$$p(x_i|\boldsymbol{\theta}) = \begin{cases} \theta^i & (1 \geq i \geq n) \\ 1 - \sum_{i=1}^n \theta^i & (i = n+1) \end{cases}$$

例えば，さいころの場合 $n = 5$，$\theta^1 = \theta^2 = \cdots = \theta^5 = 1/6$ である．この確率分布は 5 次元多様体の 1 点になる．

1.1.2　フィッシャー情報行列

統計モデル $M=\{p_{\boldsymbol{\theta}}=p(\boldsymbol{x}|\boldsymbol{\theta})\,|\,\boldsymbol{\theta}=(\theta^1,\dots,\theta^n)\in\Theta\}$ に対して，

$$g_{ij}(\boldsymbol{\theta})=E_{\boldsymbol{\theta}}\left[\partial_i l(x|\boldsymbol{\theta})\partial_j l(x|\boldsymbol{\theta})\right]=\int\frac{\partial l(x|\boldsymbol{\theta})}{\partial\theta^i}\frac{\partial l(x|\boldsymbol{\theta})}{\partial\theta^j}p(x|\boldsymbol{\theta})dx$$

をフィッシャー情報行列と呼ぶ.

$$\partial_i=\frac{\partial}{\partial\theta^i}$$
$$l(x|\boldsymbol{\theta})=\log p(x|\boldsymbol{\theta})$$
$$E_{\boldsymbol{\theta}}[f]=\int f(x)p(x|\boldsymbol{\theta})dx$$

である．また，$g_{ij}(\boldsymbol{\theta})$ を (i,j) 要素に持つ行列を $\boldsymbol{G}(\boldsymbol{\theta})$ と書く．このとき，$\boldsymbol{G}(\boldsymbol{\theta})$ は対称行列（$g_{ij}=g_{ji}$）であり，任意のベクトル $\boldsymbol{c}=(c^1,\dots,c^n)$ に対して，

$$\boldsymbol{c}^T\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{c}=\sum_{ij}c^ic^jg_{ij}(\boldsymbol{\theta})=\int\sum_{ij}c^ic^j\frac{\partial l(x|\boldsymbol{\theta})}{\partial\theta^i}\frac{\partial l(x|\boldsymbol{\theta})}{\partial\theta^j}p(x|\boldsymbol{\theta})dx=\int\sum_i\left(c^i\frac{\partial l(x|\boldsymbol{\theta})}{\partial\theta^i}\right)^2p(x|\boldsymbol{\theta})dx\geq 0$$

であるので，$\boldsymbol{G}(\boldsymbol{\theta})$ は半正定値行列である．ここでは，$\boldsymbol{G}(\boldsymbol{\theta})$ が正定値行列となる場合に話を限定する．これにより，g_{ij} はリーマン計量とみなせる．このとき，g_{ij} をフィッシャー計量と呼ぶ.

ところで，フィッシャー計量の定義式は

$$g_{ij}=\int\left(\frac{\partial}{\partial\theta^i}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta^j}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)p(\boldsymbol{x}|\boldsymbol{\theta})dx$$

である．ここで，よく使う式変形のトリック

$$\left(\frac{\partial}{\partial\theta^j}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)p(\boldsymbol{x}|\boldsymbol{\theta})=\frac{1}{p(\boldsymbol{x}|\boldsymbol{\theta})}p(\boldsymbol{x}|\boldsymbol{\theta})\left(\frac{\partial}{\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta})\right)=\frac{\partial}{\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta})$$

を用いれば，

$$g_{ij}=-\int\left(\frac{\partial^2}{\partial\theta^i\partial\theta^j}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)p(\boldsymbol{x}|\boldsymbol{\theta})dx=-E_{\boldsymbol{\theta}}\left[\partial_i\partial_j l(\boldsymbol{x}|\boldsymbol{\theta})\right]\tag{1.1}$$

を得る．証明の詳細は第 1.4.1 節に示す.

1.1.3　フィッシャー計量のもっともらしさ

n 次元統計モデル,

$$M=\{p_{\boldsymbol{\theta}}=p(\boldsymbol{x}|\boldsymbol{\theta})|\boldsymbol{\theta}=(\theta^1,\dots,\theta^n)\in\Theta\}$$

において，図 1.1 に示すように，パラメータ $\boldsymbol{\theta}$ が $\boldsymbol{\theta}\to\boldsymbol{\theta}+d\boldsymbol{\theta}$ と微小変化したとき多様体上の点 $p(\boldsymbol{x}|\boldsymbol{\theta})$ は $p(\boldsymbol{x}|\boldsymbol{\theta}+d\boldsymbol{\theta})$ と変位する．$\boldsymbol{\theta}\to\boldsymbol{\theta}+d\boldsymbol{\theta}$ にともなう，パラメータの微小変位 $d\boldsymbol{\theta}$ と，多様体上の微小変位 ds の関係を導く.

2 点 $p(\boldsymbol{x}|\boldsymbol{\theta})$，$p(\boldsymbol{x}|\boldsymbol{\theta}+d\boldsymbol{\theta})$ 間の微小距離は，KL ダイバージェンスを用いて，

$$ds^2=KL[p(\boldsymbol{x}|\boldsymbol{\theta})\|p(\boldsymbol{x}|\boldsymbol{\theta}+d\boldsymbol{\theta})]=\int p(\boldsymbol{x}|\boldsymbol{\theta})\log p(\boldsymbol{x}|\boldsymbol{\theta})d\boldsymbol{x}-\int p(\boldsymbol{x}|\boldsymbol{\theta})\log p(\boldsymbol{x}|\boldsymbol{\theta}+d\boldsymbol{\theta})d\boldsymbol{x}$$

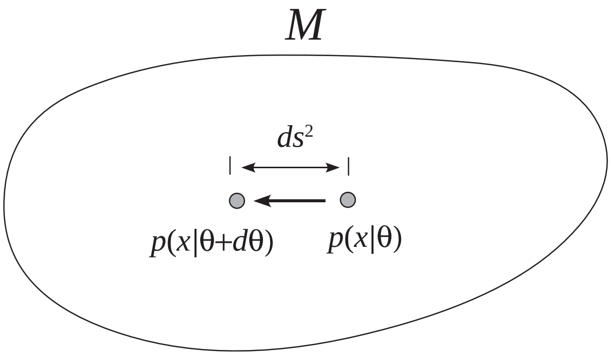

図 1.1: パラメータの微小変位 $d\boldsymbol{\theta}$ と，多様体上の微小変位 ds の関係は，KL ダイバージェンスを用いて $ds^2 = KL[p(\boldsymbol{x}|\boldsymbol{\theta})\|p(\boldsymbol{x}|\boldsymbol{\theta}+d\boldsymbol{\theta})]$ と表される．

と与えられる．右辺をテイラー展開すると，

$$ds^2 = KL[p(\boldsymbol{x}|\boldsymbol{\theta})\|p(\boldsymbol{x}|\boldsymbol{\theta})] + \left(\frac{\partial}{\partial\theta^i}KL[p(\boldsymbol{x}|\boldsymbol{\theta})\|p(\boldsymbol{x}|\boldsymbol{\theta})]\right)d\theta^i + \frac{1}{2}\left(\frac{\partial^2}{\partial\theta^i\partial\theta^j}KL[p(\boldsymbol{x}|\boldsymbol{\theta})\|p(\boldsymbol{x}|\boldsymbol{\theta})]\right)d\theta^i d\theta^j$$

である．右辺第 1 項はゼロである．第 2 項は

$$\frac{\partial}{\partial\theta^i}KL[p(\boldsymbol{x}|\boldsymbol{\theta})\|p(\boldsymbol{x}|\boldsymbol{\theta})] = \frac{\partial}{\partial\theta^i}\left[\int p(\boldsymbol{x}|\boldsymbol{\theta})\log p(\boldsymbol{x}|\boldsymbol{\theta})d\boldsymbol{x} - \int p(\boldsymbol{x}|\boldsymbol{\theta})\log p(\boldsymbol{x}|\boldsymbol{\theta})d\boldsymbol{x}\right]$$

であるが，右辺第 1 項は定数である（$\boldsymbol{\theta}$ は動かさない）．右辺第 2 項の $\log p(\boldsymbol{x}|\boldsymbol{\theta})$ の $\boldsymbol{\theta}$ が変数であるので，

$$\begin{aligned}\frac{\partial}{\partial\theta^i}KL[p(\boldsymbol{x}|\boldsymbol{\theta})\|p(\boldsymbol{x}|\boldsymbol{\theta})] &= -\int p(\boldsymbol{x}|\boldsymbol{\theta})\left(\frac{\partial}{\partial\theta^i}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)d\boldsymbol{x} = -\int p(\boldsymbol{x}|\boldsymbol{\theta})\left(\frac{1}{p(\boldsymbol{x}|\boldsymbol{\theta})}\frac{\partial}{\partial\theta^i}p(\boldsymbol{x}|\boldsymbol{\theta})\right)d\boldsymbol{x}\\ &= -\int\frac{\partial}{\partial\theta^i}p(\boldsymbol{x}|\boldsymbol{\theta})d\boldsymbol{x} = -\frac{\partial}{\partial\theta^i}\int p(\boldsymbol{x}|\boldsymbol{\theta})d\boldsymbol{x} = -\frac{\partial}{\partial\theta^i}1 = 0\end{aligned}$$

したがって，第 2 項もゼロである．第 3 項は，1.4.1 節の導出を使って，

$$\begin{aligned}\frac{\partial^2}{\partial\theta^i\partial\theta^j}KL[p(\boldsymbol{x}|\boldsymbol{\theta})\|p(\boldsymbol{x}|\boldsymbol{\theta})] &= -\int p(\boldsymbol{x}|\boldsymbol{\theta})\frac{\partial^2}{\partial\theta^i\partial\theta^j}\log p(\boldsymbol{x}|\boldsymbol{\theta})d\boldsymbol{x} = \int\left[\frac{\partial}{\partial\theta^i}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta})\right]d\boldsymbol{x}\\ = \int\left[\frac{\partial}{\partial\theta^i}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right]\left[\frac{1}{p(\boldsymbol{x}|\boldsymbol{\theta})}\frac{\partial}{\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta})\right]p(\boldsymbol{x}|\boldsymbol{\theta})d\boldsymbol{x} &= \int\left[\frac{\partial}{\partial\theta^i}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta^j}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right]p(\boldsymbol{x}|\boldsymbol{\theta})d\boldsymbol{x}\\ &= E_{\boldsymbol{\theta}}\left[\left(\frac{\partial}{\partial\theta^i}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta^j}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)\right] = g_{ij}(\boldsymbol{\theta})\end{aligned}$$

であるので，結局，

$$ds^2 = \frac{1}{2}g_{ij}(\boldsymbol{\theta})d\theta^i d\theta^j$$

を得る．つまり，フィシャー行列を計量とすれば，パラメータ $\boldsymbol{\theta}$ の微小変化 $d\theta^i$ に対応する，多様体上の位置の微小変位 ds^2 を（1/2 倍を無視して）

$$ds^2 = g_{ij}d\theta^i d\theta^j$$

と第 1 基本形式の形で表すことができる．フィシャー行列は正定値行列であるので[1]，フィシャー行列を計量とすることにより統計モデル M はリーマン多様体となる．

[1]実際には，非負定置（$G \succeq 0$）なので，正定値（$G \succ 0$）になる場合に話を限定する．

1.1.4　1 次元正規分布の多様体

$$\mathcal{N}(x\,|\,\mu,\sigma) = p(x\,|\,\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

であり，

$$\log p(x\,|\,\mu,\sigma) = -\log\sqrt{2\pi} - \log\sigma - \frac{(x-\mu)^2}{2\sigma^2}$$

である．多様体 M は

$$M = \{p(x\,|\,\mu,\sigma)\,|\,\mu\in\mathbb{R},\sigma\in\mathbb{R}^+\}$$

と表せる．ここで，$\mathbb{R}^+$ はゼロを含む正の実数領域を表す.

フィッシャー計量を用いるのに式 (1.1) を用いる．そのため，まず,

$$\frac{\partial^2\log p(x\,|\,\mu,\sigma)}{\partial\mu^2} = -\frac{1}{\sigma^2}$$

$$\frac{\partial^2\log p(x\,|\,\mu,\sigma)}{\partial\mu\partial\sigma} = -2\frac{x-\mu}{\sigma^3}$$

$$\frac{\partial^2\log p(x\,|\,\mu,\sigma)}{\partial\sigma^2} = \frac{1}{\sigma^2} - 3\frac{(x-\mu)^2}{\sigma^4}$$

を計算する．これらを用いて,

$$g_{11} = -E_{\boldsymbol{\theta}}\left[-\frac{1}{\sigma^2}\right] = \frac{1}{\sigma^2}$$

$$g_{12} = g_{21} = -E_{\boldsymbol{\theta}}\left[-2\frac{x-\mu}{\sigma^3}\right] = 0$$

$$g_{22} = -E_{\boldsymbol{\theta}}\left[\frac{1}{\sigma^2} - 3\frac{(x-\mu)^2}{\sigma^4}\right] = -\left(\frac{1}{\sigma^2} - 3\frac{E_{\boldsymbol{\theta}}[(x-\mu)^2]}{\sigma^4}\right) = \frac{2}{\sigma^2}$$

したがって，$\theta^1=\mu$，$\theta^2=\sigma$ とすれば,

$$\boldsymbol{G}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix}$$

を得る．したがって，M の第 1 基本形式は

$$ds^2 = \frac{1}{\sigma^2}[d\mu,\,d\sigma]\begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix}\begin{bmatrix}d\mu\\ d\sigma\end{bmatrix} = \frac{d\mu^2+2d\sigma^2}{\sigma^2} \tag{1.2}$$

を得る．さらに座標変換 $(\mu,\sigma)\to(x,y)=(\frac{\mu}{\sqrt{2}},\sigma)$ を行い，この (x,y) で計量を表せば,

$$ds^2 = \frac{dx^2+dy^2}{y^2} \tag{1.3}$$

と表せる.

領域 $\{(x,y);y>0\}$ に対して定義された式 (1.3) の計量はポアンカレの上半面モデルとかポアンカレ計量と呼ばれており，ロバチェフスキーの双曲線幾何学として良く知られているものである.

1.2　正規分布間の距離の計算

1.2.1　ユークリッド距離

スカラー正規分布多様体 M の 2 点 $p_1(x|(\mu_1,\sigma_1))$ と $p_2(x|(\mu_2,\sigma_2))$ のユークリッド距離 d_E は,

$$d_E^2 = (\mu_1-\mu_2)^2 + (\sigma_1-\sigma_2)^2$$

である．しかし，ピーク中心の隔たりとピークの広がり方の隔たりを同じに取り扱っていて，合理的とは思えず.

1.2.2　ポアンカレ計量における測地線

測地線方程式

2 つの正規分布が与えられたとき，その間の距離を計算してみる．そのために，ポアンカレ計量における測地線の方程式を導く．この計量は

$$g_{11} = \frac{1}{y^2}, \quad g_{12} = g_{21} = 0, \quad g_{22} = \frac{1}{y^2}$$

と表される．1.4.2 節の議論により，ゼロでないクリストッフェル記号は

$$\Gamma^1{}_{12} = \Gamma^1{}_{21} = \Gamma^2{}_{22} = -\frac{1}{y} \qquad \Gamma^2{}_{11} = \frac{1}{y}$$

である．これらを測地線の方程式

$$\frac{\partial^2 x^1}{\partial s^2} + \Gamma^1{}_{ij}\frac{dx^i}{ds}\frac{dx^j}{ds} = 0$$
$$\frac{\partial^2 x^2}{\partial s^2} + \Gamma^2{}_{ij}\frac{dx^i}{ds}\frac{dx^j}{ds} = 0$$

に代入すれば，弧長パラメータ s を用いて,（さらに，$x^1 \to x$, $x^2 \to y$ と書き直して）測地線方程式

$$\frac{\partial^2 x}{\partial s^2} - \frac{2}{y}\frac{dx}{ds}\frac{dy}{ds} = 0 \tag{1.4}$$

$$\frac{\partial^2 y}{\partial s^2} + \frac{1}{y}\left((\frac{dx}{ds})^2 - (\frac{dy}{ds})^2\right) = 0 \tag{1.5}$$

を得る．

測地線方程式の解

連立微分方程式（(1.4) と (1.5)）を解いて解を求める．

まず，式 (1.4) から，$y \neq 0$ より，s での微分を x' や y' と表し，

$$\frac{x''}{y^2} - \frac{2x'y'}{y^3} = 0$$

を得る．これは,

$$\left(\frac{x'}{y^2}\right)' = 0$$

に等しいので,

$$x' = ay^2$$

を得る．ここで，a は積分定数である．$a = 0$ の場合,

$$x' = 0 \quad \rightarrow \quad x = c$$

と言う解を得る．すなわち，$a = 0$ に対応する解として y 軸に平行な直線を得る．

次に，式 (1.5) から,

$$\frac{y''}{y} - \frac{(y')^2}{y^2} + \frac{(x')^2}{y^2} = 0 \quad \rightarrow \quad \left(\frac{y'}{y}\right)' + \frac{(x')^2}{y^2} = 0$$

を導くことができる．これに，$x' = ay^2$ を代入すると,

$$\left(\frac{y'}{y}\right)' + \frac{x'ay^2}{y^2} = 0 \quad \rightarrow \quad \left(\frac{y'}{y}\right)' + ax' = 0$$

であり，積分を実行すると

$$\frac{y'}{y} + ax = b \quad \rightarrow \quad y' = (-ax + b)y$$

が得られる．ここで,

$$x' = ay^2$$
$$y' = (-ax + b)y$$

を連立させると，$a \neq 0$ として,

$$\frac{y'}{x'} = \frac{-ax + b}{ay} \quad \rightarrow \quad aydy = -axdx + bdx$$

さらに積分を実行すると,

$$\frac{a}{2}y^2 = -\frac{a}{2}x^2 + bx + C$$

を得，最終的に

$$y^2 + (x - b')^2 = C'$$

を得る．上式は，x 軸上に中心を持つ任意の半径の半円が測地線になることを示している．

まとめると，半平面 $\{(x, y) \,|\, (x, y) \in \mathbb{R}^2, y > 0\}$ において定義された第 1 基本形式

$$ds^2 = \frac{dx^2 + dy^2}{y^2}$$

の測地線は，x 軸上に中心を持つ任意半径の半円と x 軸に垂直な直線である．この計量の定義された半平面はポアンカレ半平面と呼ばれ $\mathcal{H}$ で表す．

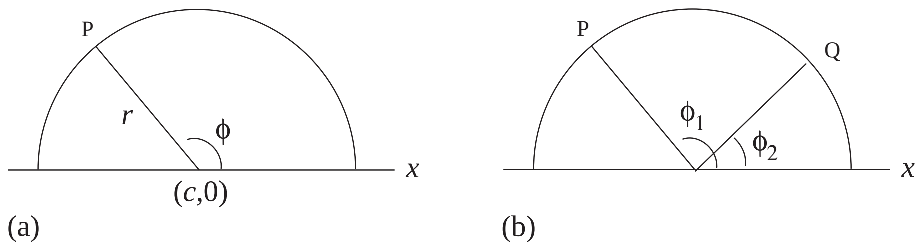

図 1.2:

1.2.3 ポアンカレ半平面 $\mathcal{H}$ における 2 点間の距離

簡単な導出

ポアンカレ計量の定義された半平面 $\mathcal{H}$ に 2 つの点 P と Q を仮定する．2 点 P，Q を通る測地線が半円の場合，中心を $(c, 0)$ として半径を r とする．図 1.2 (a) に示すごとく，半円上の点が x 軸正方向となす角度を ϕ とする．さらに図 1.2 (b) に示すごとく，P と Q に対するこの角度を ϕ_1，ϕ_2 とすれば，

$$x = c + r\cos\phi$$
$$y = r\sin\phi$$

$\mathcal{H}$ の P, Q 間の距離 $d_H(P, Q)$ は，

$$d_H(P,Q) = \int_P^Q ds = \int_P^Q \frac{\sqrt{dx^2 + dy^2}}{y}$$

で与えられるので，

$$dx = -r\sin\phi d\phi \qquad dy = r\cos\phi d\phi$$

を代入すれば，距離 $d_H(P, Q)$ は

$$d_H(P,Q) = \int_P^Q \frac{\sqrt{(r\sin\phi)^2 + (r\cos\phi)^2}}{r\sin\phi} d\phi = \int_{\phi_1}^{\phi_2} \frac{1}{\sin\phi} d\phi = \left|\log \frac{\tan(\phi_2/2)}{\tan(\phi_1/2)}\right| \tag{1.6}$$

ところで，P, Q を結ぶ測地線が y 軸と平行な場合は，

$$d_H(P,Q) = \int_P^Q ds = \int_P^Q \frac{\sqrt{dx^2 + dy^2}}{y} = \int_P^Q \frac{\sqrt{dy^2}}{y} = \int_P^Q \frac{dy}{y}$$

であるので，P, Q の y 座標をそれぞれ a, b とすれば，

$$d_H(P,Q) = \left|[\log y]_a^b\right| = \left|\log \frac{b}{a}\right| \tag{1.7}$$

を得る．

複素数を用いた導出

上の議論では PQ 間距離を，測地線が半円の場合と y 軸に平行な直線の場合で分けて考え式 (1.6)，(1.7) を得たが，この 2 つの式は，1 つの式，すなわち，

$$d_H(P,Q) = \left|\log \frac{1+\Lambda}{1-\Lambda}\right| \qquad \Lambda = \left|\frac{z-w}{z-\bar{w}}\right| \tag{1.8}$$

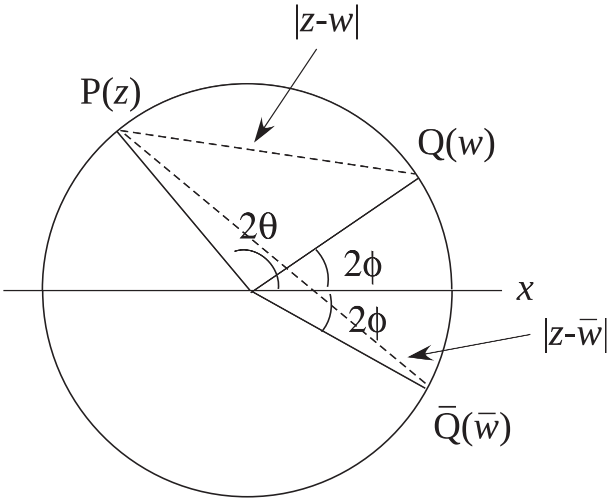

図 1.3:

と表すことができる．ここで，z と w は，P, Q の半平面上の位置 (x_1, y_1), (x_2, y_2) を複素数として表した，$z = x_1 + iy_1, w = x_2 + iy_2$ である．また，$\bar{w} = x_2 - iy_2$ である．

証明： 図 1.3 に示すように点 P と Q が x 軸正方向となす角を 2θ と 2ϕ とする．(1.2.3 節の議論では，これらは ϕ_1 と ϕ_2 としていた．) すると，図 1.3 からわかるように，

$$|z - w| = 2\sin(\theta - \phi) \qquad |z - \bar{w}| = 2\sin(\theta + \phi)$$

である．したがって，

$$\frac{1+\Lambda}{1-\Lambda} = \frac{\sin(\theta+\phi)+\sin(\theta+\phi)}{\sin(\theta+\phi)-\sin(\theta+\phi)} = \frac{\tan\theta}{\tan\phi}$$

となり，

$$\left|\log\frac{1+\Lambda}{1-\Lambda}\right| = \left|\log\frac{\tan\theta}{\tan\phi}\right| = d(P, Q)$$

である．また，PQ の実部が等しい場合には，

$$|z - w| = |y_1 - y_2| \quad \text{および} \quad |z - \bar{w}| = |y_1 + y_2|, \quad \text{すなわち} \quad \frac{1+\Lambda}{1-\Lambda} = \frac{y_1}{y_2}$$

を得る．したがって，この場合も式 (1.8) により $d_H(P, Q)$ を表すことができる．

1.2.4 $\mathcal{H}$ における距離から正規分布間の距離へ

2 つの正規分布 $\mathcal{N}(x|\mu_1, \sigma_1)$ と $\mathcal{N}(x|\mu_2, \sigma_2)$ の距離をもとめる．正規分布多様体 N における点 (μ_1, σ_1) と点 (μ_2, σ_2) 間の距離 d_N は，ポアンカレ半平面 $\mathcal{H}$ 上の点 $P(\frac{\mu_1}{\sqrt{2}}, \sigma_1)$ と $Q(\frac{\mu_2}{\sqrt{2}}, \sigma_2)$ 間の距離 $d_H(P, Q)$ を用いて，

$$d_N\left((\mu_1, \sigma_1), (\mu_2, \sigma_2)\right) = \sqrt{2}d_H(P, Q) = \sqrt{2}d_H\left((\frac{\mu_1}{\sqrt{2}}, \sigma_1), (\frac{\mu_2}{\sqrt{2}}, \sigma_2)\right) \tag{1.9}$$

として計算できる．ここで，$d_H(P, Q)$ は式 (1.8)

$$d_H(P, Q) = \left|\log\frac{1+\Lambda}{1-\Lambda}\right| \qquad \Lambda = \left|\frac{z - w}{z - \bar{w}}\right|$$

において，

$$z = \frac{\mu_1}{\sqrt{2}} + i\sigma_1 \qquad w = \frac{\mu_2}{\sqrt{2}} + i\sigma_2 \qquad \bar{w} = \frac{\mu_2}{\sqrt{2}} - i\sigma_2$$

として計算する．

例1

2 つの正規分布 $\mathcal{N}(x|\mu_1, \sigma_1)$，$\mathcal{N}(x|\mu_2, \sigma_2)$ において，$\mu_1 = \mu_2 = \mu$ である場合（$\sigma_1 > \sigma_2$ を仮定）：

$$z = \frac{\mu}{\sqrt{2}} + i\sigma_1 \qquad w = \frac{\mu}{\sqrt{2}} + i\sigma_2 \qquad \bar{w} = \frac{\mu}{\sqrt{2}} - i\sigma_2$$

であり，

$$\Lambda = \left|\frac{z-w}{z-\bar{w}}\right| = \frac{\sigma_1 - \sigma_2}{\sigma_1 + \sigma_2} \quad \rightarrow \quad d_H(z,w) = \log\frac{1+\Lambda}{1-\Lambda} = \log\frac{\sigma_1}{\sigma_2}$$

である．結局

$$d_N\left((\mu_1, \sigma_1), (\mu_2, \sigma_2)\right) = \sqrt{2} d_H(z,w) = \sqrt{2}\log\frac{\sigma_1}{\sigma_2}$$

となる．

例2

2 つの正規分布 $\mathcal{N}(x|\mu_1, \sigma_1)$，$\mathcal{N}(x|\mu_2, \sigma_2)$ において，$\sigma_1 = \sigma_2 = \sigma$ である場合（$\mu_1 > \mu_2$ を仮定）：

$$z = \frac{\mu_1}{\sqrt{2}} + i\sigma \qquad w = \frac{\mu_2}{\sqrt{2}} + i\sigma \qquad \bar{w} = \frac{\mu_2}{\sqrt{2}} - i\sigma$$

であり，

$$\Lambda = \left|\frac{z-w}{z-\bar{w}}\right| = \frac{\mu_1 - \mu_2}{\mu_1 - \mu_2 + 2\sqrt{2}\sigma} = \frac{\mathit{\Delta}\mu}{\mathit{\Delta}\mu + 2\sqrt{2}\sigma} \qquad (\mathit{\Delta}\mu = \mu_1 - \mu_2)$$

であるので，

$$d_H(z,w) = \log\frac{1+\Lambda}{1-\Lambda} = \log\frac{\mathit{\Delta}\mu + \sqrt{2}\sigma}{\sqrt{2}\sigma} = \log\left(1 + \frac{\mathit{\Delta}\mu}{\sqrt{2}\sigma}\right)$$

すなわち，

$$d_N\left((\mu_1, \sigma), (\mu_2, \sigma)\right) = \sqrt{2} d_H(z,w) = \sqrt{2}\log\left(1 + \frac{\mathit{\Delta}\mu}{\sqrt{2}\sigma}\right)$$

を得る．上式は，判別分析などで経験的に使われるマハラノビス距離に近いものである．

1.3　曲率の計算

正規分布多様体のリーマン曲率を計算してみる．リーマン曲率テンソルは

$$R^i_{jkl} = \frac{\partial \Gamma^i{}_{jl}}{\partial x^k} - \frac{\partial \Gamma^i{}_{jk}}{\partial x^l} + \Gamma^r{}_{jl}\,\Gamma^i{}_{rk} - \Gamma^r{}_{jk}\,\Gamma^i{}_{rl}$$

である．添え字に対する対称性のため，$n = 2$ の場合でゼロでない値を持つのは R^1_{212} とこれの対称形のみである．R^1_{212} を求めてみると，

$$R^1_{212} = \frac{\partial \Gamma^1{}_{22}}{\partial x^1} - \frac{\partial \Gamma^1{}_{21}}{\partial x^2} + \Gamma^r{}_{22}\,\Gamma^1{}_{r1} - \Gamma^r{}_{21}\,\Gamma^1{}_{r2}$$

ポアンカレ半平面の曲率

まず，1.4.2 節の結果を用いて，$\mathcal{H}$ の場合で $R^1_{\ 212}$ を求めてみると，

$$\frac{\partial \Gamma^1{}_{22}}{\partial x^1} = 0 \qquad \text{なぜなら} \quad \Gamma^1{}_{22} = 0$$

$$\frac{\partial \Gamma^1{}_{21}}{\partial x^2} = \frac{\partial}{\partial y}(-\frac{1}{y}) = \frac{1}{y^2}$$

$$\Gamma^r{}_{22}\,\Gamma^1{}_{r1} = \Gamma^1{}_{22}\,\Gamma^1{}_{11} + \Gamma^2{}_{22}\,\Gamma^1{}_{21} = \Gamma^2{}_{22}\,\Gamma^1{}_{21} = \frac{1}{y^2}$$

$$\Gamma^r{}_{21}\,\Gamma^1{}_{r2} = \Gamma^1{}_{21}\,\Gamma^1{}_{12} + \Gamma^2{}_{21}\,\Gamma^1{}_{22} = \Gamma^1{}_{21}\,\Gamma^1{}_{12} = \frac{1}{y^2}$$

であり，まとめて，

$$R^1_{\ 212} = 0 - \frac{1}{y^2} + \frac{1}{y^2} - \frac{1}{y^2} = -\frac{1}{y^2}$$

を得る．対応する第 1 種リーマンテンソル R_{1212} は

$$R_{1212} = g_{1\ell} R^\ell_{\ 212} = g_{11} R^1_{\ 212} = \frac{1}{y^2}(-\frac{1}{y^2}) = -\frac{1}{y^4}$$

となる．R_{1212} を用いるとガウス曲率 K は,

$$K = \frac{R_{1212}}{g_{11}g_{22} - g_{12}^2} = -1 \qquad \text{なぜなら} \quad g_{11}g_{22} - g_{12}^2 = \frac{1}{y^4}$$

を得る．ポアンカレ半平面 $\mathcal{H}$ において，ガウス曲率がいたるところ -1 であるのは良く知られた結果である．

1 次元正規分布多様体の曲率

正規分布多様体 N の場合とポアンカレ半平面 $\mathcal{H}$ の場合とで，ことなる接続係数は $\Gamma^2{}_{11}$ のみである．しかしながら，$\mathcal{H}$ に対する $R^1_{\ 212}$ の計算に $\Gamma^2{}_{11}$ は使われていないので，$\mathcal{H}$ に対する $R^1_{\ 212}$ の計算結果はそのまま正規分布多様体 N の場合の $R^1_{\ 212}$ とすることができる．また,

$$R_{1212} = g_{1\ell} R^\ell_{\ 212} = g_{11} R^1_{\ 212}$$

であるので，R_{1212} の計算結果はそのまま正規分布多様体の場合の結果となる．しかしながら，ガウス曲率は

$$K = \frac{R_{1212}}{g_{11}g_{22} - g_{12}^2} = -\frac{1}{2} \qquad \text{なぜなら} \quad g_{11}g_{22} - g_{12}^2 = \frac{2}{y^4}$$

となる．ポアンカレ半平面の場合と同様，定曲率空間であるが，ガウス曲率の値は異なる．

1.4 Appendix

1.4.1 式 (1.1) の証明

式 (1.1) を示す.（以下の証明では確率変数での積分を $\sum_{\boldsymbol{x}\in\Omega}$ と表記している.）を示す．逆からスタートすれば

$$\begin{aligned}\sum_{\boldsymbol{x}\in\Omega}\left(\frac{\partial^2}{\partial\theta^i\partial\theta^j}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)p(\boldsymbol{x}|\boldsymbol{\theta}) &= \sum_{\boldsymbol{x}\in\Omega}\left[\frac{\partial}{\partial\theta^i}\left(\frac{\partial}{\partial\theta^j}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)\right]p(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\Omega}\left[\frac{\partial}{\partial\theta^i}\frac{\frac{\partial}{\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x}|\boldsymbol{\theta})}\right]p(\boldsymbol{x}|\boldsymbol{\theta}) \\ &= \sum_{\boldsymbol{x}\in\Omega}\left[\frac{\left[\frac{\partial^2}{\partial\theta^i\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta})\right]p(\boldsymbol{x}|\boldsymbol{\theta}) - \left[\frac{\partial}{\partial\theta^i}p(\boldsymbol{x}|\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta})\right]}{p(\boldsymbol{x}|\boldsymbol{\theta})^2}\right]p(\boldsymbol{x}|\boldsymbol{\theta}) \\ &= \sum_{\boldsymbol{x}\in\Omega}\frac{\partial^2}{\partial\theta^i\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta}) - \sum_{\boldsymbol{x}\in\Omega}\left[\frac{\partial}{\partial\theta^i}p(\boldsymbol{x}|\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta})\right]\frac{1}{p(\boldsymbol{x}|\boldsymbol{\theta})}\end{aligned}$$

上式の右辺第 1 項は,

$$\sum_{\boldsymbol{x}\in\Omega}\frac{\partial^2}{\partial\theta^i\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{\partial^2}{\partial\theta^i\partial\theta^j}\sum_{\boldsymbol{x}\in\Omega}p(\boldsymbol{x}|\boldsymbol{\theta}) = 0$$

である．また，第 2 項は

$$\begin{aligned}-\sum_{\boldsymbol{x}\in\Omega}\left[\frac{\partial}{\partial\theta^i}p(\boldsymbol{x}|\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta})\right]\frac{1}{p(\boldsymbol{x}|\boldsymbol{\theta})} &= -\sum_{\boldsymbol{x}\in\Omega}\left[\frac{1}{p(\boldsymbol{x}|\boldsymbol{\theta})}\frac{\partial}{\partial\theta^i}p(\boldsymbol{x}|\boldsymbol{\theta})\right]\left[\frac{1}{p(\boldsymbol{x}|\boldsymbol{\theta})}\frac{\partial}{\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta})\right]p(\boldsymbol{x}|\boldsymbol{\theta}) \\ &= -\sum_{\boldsymbol{x}\in\Omega}\left[\frac{\partial}{\partial\theta^i}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta^j}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right]p(\boldsymbol{x}|\boldsymbol{\theta})\end{aligned}$$

であるので,

$$\begin{aligned}\sum_{\boldsymbol{x}\in\Omega}\left(\frac{\partial^2}{\partial\theta^i\partial\theta^j}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)p(\boldsymbol{x}|\boldsymbol{\theta}) &= -\sum_{\boldsymbol{x}\in\Omega}\left(\frac{\partial}{\partial\theta^i}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta^j}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)p(\boldsymbol{x}|\boldsymbol{\theta}) \\ &= -E_{\boldsymbol{\theta}}\left[\left(\frac{\partial}{\partial\theta^i}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta^j}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)\right]\end{aligned}$$

が成り立つ．すなわち，式 (1.1) を示すことができた.

1.4.2 ポアンカレの上半平面計量における接続係数の計算

クリストッフェル記号は，計量 g_{ij} が与えられた場合，良く知られた公式

$$\Gamma^{\ell}{}_{ij} = \frac{1}{2}g^{\ell k}\left(\frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ki}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k}\right) \tag{1.10}$$

から計算できる[2]．ちなみに，この式で与えられる接続係数はリーマン接続の係数であり，捩率ゼロ $\Gamma^{\ell}{}_{ij} = \Gamma^{\ell}{}_{ji}$ を満たし，また，内積に対して方向微分のライプニッツ則

$$\boldsymbol{X}g(\boldsymbol{Y},\boldsymbol{Z}) = g(\nabla_X\boldsymbol{Y},\boldsymbol{Z}) + g(\boldsymbol{Y},\nabla_X\boldsymbol{Z})$$

[2] この式はおぼえにくい．簡単に求めるには，まず，第 1 種クリストッフェル記号を

$$\Gamma_k ij = \frac{1}{2}\left(-g_{ij,k} + g_{jk,i} + g_{ki,j}\right)$$

を満たす．したがって，この式を用いて得られた接続係数は自己双対 $\nabla^* = \nabla$ である．

クリストッフェル記号の多くがゼロである場合，式 (1.10) からクリストッフェル記号を計算するのは見通しが悪い．(ゼロであるクリストッフェル記号もすべて計算しなければならないため．) g_{ij} が対角計量である場合には，以下の公式が便利である．

添え字 α とに対して，

$$\Gamma^{\alpha}{}_{\alpha\beta} = \Gamma^{\alpha}{}_{\beta\alpha} = \frac{\partial}{\partial x^{\beta}}\left(\frac{1}{2}\log|g_{\alpha\alpha}|\right) \tag{1.11}$$

$$\Gamma^{\alpha}{}_{\beta\beta} = -\frac{1}{2g_{\alpha\alpha}}\frac{\partial g_{\beta\beta}}{\partial x^{\alpha}} \qquad (\beta \neq \alpha) \tag{1.12}$$

$$\Gamma^{k}{}_{ij} = 0 \quad \text{上記以外} \tag{1.13}$$

である．計量は

$$g_{11} = \frac{1}{(x^2)^2}, \quad g_{12} = g_{21} = 0, \quad g_{22} = \frac{1}{(x^2)^2}$$

である．式 (1.11) に対して，$\beta = 2, \alpha = 1$ がゼロでない値を持つ場合である．

$$\Gamma^{1}{}_{12} = \Gamma^{1}{}_{21} = \frac{\partial}{\partial x^2}\left(\frac{1}{2}\log|g_{11}|\right) = \frac{1}{2}\frac{\partial}{\partial x^2}\log(\frac{1}{(x^2)^2}) = -\frac{1}{(x^2)} = -\frac{1}{y}$$

さらに，$\beta = 2, \alpha = 2$ もゼロでない値を持つ

$$\Gamma^{2}{}_{22} = \frac{1}{2}\frac{\partial}{\partial x^2}\log(\frac{1}{(x^2)^2}) = -\frac{1}{(x^2)} = -\frac{1}{y}$$

また，(1.12) に対しては，$\Gamma^{1}{}_{22}$, $\Gamma^{2}{}_{11}$ の可能性があるが，ゼロでないのは

$$\Gamma^{2}{}_{11} = -\frac{1}{2g_{22}}\frac{\partial g_{11}}{\partial x^2} = -\frac{1}{2}(x^2)^2(-\frac{2}{(x^2)^3}) = \frac{1}{x^2} = \frac{1}{y}$$

のみである．

残りのクリストッフェル記号はゼロである．

1.4.3　正規分布多様体でのリーマン接続係数の計算

次は，式 (1.2) の正規分布多様体で接続係数を計算する．計量は

$$g_{11} = \frac{1}{(x^2)^2}, \quad g_{12} = g_{21} = 0, \quad g_{22} = \frac{2}{(x^2)^2}$$

である．計量の変数依存性は $\mathcal{H}$ の場合と同じであるので，式 (1.11) に対して，$\beta = 2, \alpha = 1$ がゼロでない値を持つ場合である．

$$\Gamma^{1}{}_{12} = \Gamma^{1}{}_{21} = \frac{\partial}{\partial x^2}\left(\frac{1}{2}\log|g_{11}|\right) = \frac{1}{2}\frac{\partial}{\partial x^2}\log(\frac{1}{(x^2)^2}) = -\frac{1}{(x^2)} = -\frac{1}{y}$$

から計算する．ここで，$g_{ij,k} = \partial g_{ij}/\partial x^k$ の意味である．$\Gamma_k ij$ を計算したのち，

$$\Gamma^{\ell}{}_{ij} = g^{k\ell}\Gamma_k ij$$

から接続係数を求める．

さらに，$\beta = 2, \alpha = 2$ もゼロでない値を持つ

$$\Gamma^2{}_{22} = \frac{1}{2}\frac{\partial}{\partial x^2}\log(|g_{22}|) = \frac{1}{2}\frac{\partial}{\partial x^2}\log(\frac{2}{(x^2)^2}) = -\frac{1}{(x^2)} = -\frac{1}{y}$$

また，(1.12) に対しては，$\Gamma^1{}_{22}$, $\Gamma^2{}_{11}$ の可能性があるが，ゼロでないのは

$$\Gamma^2{}_{11} = -\frac{1}{2g_{22}}\frac{\partial g_{11}}{\partial x^2} = -\frac{(x^2)^2}{4}(-\frac{2}{(x^2)^3}) = \frac{1}{2x^2} = \frac{1}{2y}$$

のみである．

残りのクリストッフェル記号はゼロである．また，ポアンカレ半平面の場合と比較して，異なるのは $\Gamma^2{}_{11} = \frac{1}{2y}$ となる場合のみである．

第2章　共変微分とベクトルの平行移動

抽象的な曲面（多様体）上での関数の微分–共変微分–を定義する．方針は以下の通りである．まず，3 次元ユークリッド空間内の曲面で，曲面に沿った微分として共変微分を定義し，その共変微分が満たす性質を調べる．そして多様体上でそのような性質を持つ写像として共変微分を定義するのである．3 次元ユークリッド空間内における共変微分が自然に持つ 5 つの性質を課すと多様体上でも共変微分が 1 つに決まり，リーマン接続と呼ばれる.（「接続」と「共変微分」は全く同じ意味で用いられる.）さらに，ベクトルの平行移動についても議論する．

2.1　ユークリッド空間におけるベクトル場の微分

2.1.1　接ベクトル方向の微分

まず，E^3 に埋め込まれた 2 次元曲面上でのベクトル場（ベクトル値関数）の微分を考える[1]．次節で，これを抽象的な曲面（多様体）へ一般化する．

まず，E^3 に埋め込まれた 2 次元曲面 S を

$$\boldsymbol{x} = (x(u^1, u^2), y(u^1, u^2), z(u^1, u^2))$$

と，E^3 における 3 次元座標で曲面を表す．すると，

$$\frac{\partial \boldsymbol{x}}{\partial u^1} = \left(\frac{\partial x}{\partial u^1}, \frac{\partial y}{\partial u^1}, \frac{\partial z}{\partial u^1}\right), \quad \frac{\partial \boldsymbol{x}}{\partial u^2} = \left(\frac{\partial x}{\partial u^2}, \frac{\partial y}{\partial u^2}, \frac{\partial z}{\partial u^2}\right)$$

である．$\dfrac{\partial \boldsymbol{x}}{\partial u^1}$ と $\dfrac{\partial \boldsymbol{x}}{\partial u^2}$ は 1 次独立であると仮定する.（正則曲面と呼ばれ，このとき曲面 S は滑らかな曲面となる.）$\boldsymbol{X}$ を点 p において S に接するベクトルとすれば，

$$\boldsymbol{X} = \xi_1 \frac{\partial \boldsymbol{x}}{\partial u^1}(p) + \xi_2 \frac{\partial \boldsymbol{x}}{\partial u^2}(p) \tag{2.1}$$

である．つまり，p において S に接する任意のベクトルは，基底 $\dfrac{\partial \boldsymbol{x}}{\partial u^1}(p)$ と $\dfrac{\partial \boldsymbol{x}}{\partial u^2}(p)$ の線形結合で表される．

基底を改めて $\boldsymbol{e}_i$ と書くことにすれば，接ベクトルは

$$\boldsymbol{X}(p) = \xi_1(p)\boldsymbol{e}_i(p) + \xi_2(p)\boldsymbol{e}_2(p)$$

と書ける．これは空間に定義されたベクトル値関数である，今，空間に定義されたスカラー関数 $f(p)$ がベクトル $\boldsymbol{X}(p)$ の方向にどう変化するかは，

$$\boldsymbol{X} f(p) = \xi_1(p)\partial_1 f(p) + \xi_2(p)\partial_2 f(p) = \xi_1(p)\frac{\partial}{\partial u^1} f(p) + \xi_2(p)\frac{\partial}{\partial u^2} f(p)$$

[1] 曲面上のベクトル場とは，曲面の各点にベクトル値が割り当てられている状態を言う．

と表すことができる．左辺の $\boldsymbol{X}f(p)$ はベクトル場 $\boldsymbol{X}$ が関数 $f(p)$ に作用して $\boldsymbol{X}$ 方向の方向微分を与えると考えている．つまり，ここで，$\boldsymbol{X}$ を接ベクトル方向の微分オペレーターと同一視している．(少し唐突であるがこのように考え，以後は $\boldsymbol{X}$ は微分オペレータを表すとする.）すなわち，$\boldsymbol{X}$ は

$$\boldsymbol{X} = \xi_1(p)\partial_1 + \xi_2(p)\partial_2 = \xi_1(p)\frac{\partial}{\partial u^1} + \xi_2(p)\frac{\partial}{\partial u^2} \tag{2.2}$$

と表される．この微分オペレーターをスカラー関数 f に作用させた場合,

$$\boldsymbol{X}f = \xi_1\frac{\partial f}{\partial u^1}(p) + \xi_2\frac{\partial f}{\partial u^2}(p)$$

となる．

2.1.2　曲面上のベクトル場の微分

$\boldsymbol{Y}(u^1, u^2)$ が S 上のベクトル値関数

$$\boldsymbol{Y}(u^1, u^2) = (Y^1(u^1, u^2), Y^2(u^1, u^2), Y^3(u^1, u^2))$$

である場合，ベクトル値関数に対するオペレーター $D_{\boldsymbol{X}}$ を,

$$D_{\boldsymbol{X}}\boldsymbol{Y} = (\boldsymbol{X}Y^1(u^1, u^2), \boldsymbol{X}Y^2(u^1, u^2), \boldsymbol{X}Y^3(u^1, u^2))$$

と定義する．$D_{\boldsymbol{X}}$ はベクトル $\boldsymbol{Y}$ に作用し，$\boldsymbol{Y}$ の $\boldsymbol{X}$ 方向の方向微分と呼ぶ[2].

また，u^i 方向の方向微分を $\boldsymbol{X}_i$ と書けば，$\boldsymbol{X}_1 = \dfrac{\partial \boldsymbol{x}}{\partial u^1}, \boldsymbol{X}_2 = \dfrac{\partial \boldsymbol{x}}{\partial u^2}$ であるので，微分オペレータとしては

$$\boldsymbol{X}_1 = \frac{\partial}{\partial u^1} \quad \boldsymbol{X}_2 = \frac{\partial}{\partial u^2}$$

である．したがって,

$$D_{\boldsymbol{X}_1}\boldsymbol{X}_1 = D_{\boldsymbol{X}_1}\frac{\partial \boldsymbol{x}}{\partial u^1} = (\boldsymbol{X}_1\frac{\partial x}{\partial u^1}, \boldsymbol{X}_1\frac{\partial y}{\partial u^1}, \boldsymbol{X}_1\frac{\partial z}{\partial u^1}) = (\frac{\partial^2 x}{\partial u^1\partial u^1}, \frac{\partial^2 y}{\partial u^1\partial u^1}, \frac{\partial^2 z}{\partial u^1\partial u^1})$$

および

$$D_{\boldsymbol{X}_2}\boldsymbol{X}_1 = D_{\boldsymbol{X}_2}\frac{\partial \boldsymbol{x}}{\partial u^1} = (\boldsymbol{X}_2\frac{\partial x}{\partial u^1}, \boldsymbol{X}_2\frac{\partial y}{\partial u^1}, \boldsymbol{X}_2\frac{\partial z}{\partial u^1}) = (\frac{\partial^2 x}{\partial u^2\partial u^1}, \frac{\partial^2 y}{\partial u^2\partial u^1}, \frac{\partial^2 z}{\partial u^2\partial u^1})$$

である．なお，これらは接ベクトルを微分作用素と同一視する議論では，しばしば

$$D_{\frac{\partial}{\partial u^1}}\frac{\partial}{\partial u^1} = (\frac{\partial^2}{\partial(u^1)^2}, \frac{\partial^2}{\partial(u^1)^2}, \frac{\partial^2}{\partial(u^1)^2})$$

および

$$D_{\frac{\partial}{\partial u^2}}\frac{\partial}{\partial u^1} = (\frac{\partial^2}{\partial u^2\partial u^1}, \frac{\partial^2}{\partial u^2\partial u^1}, \frac{\partial^2}{\partial u^2\partial u^1})$$

と書かれる．このノートもこの書き方を以後用いる．

---

[2]
$$D_{\boldsymbol{X}}\boldsymbol{A} = (XA_1, XA_2, XA_3) = (\xi_1\frac{\partial A_1}{\partial u^1} + \xi_2\frac{\partial A_1}{\partial u^2}, \xi_1\frac{\partial A_2}{\partial u^1} + \xi_2\frac{\partial A_2}{\partial u^2}, \xi_1\frac{\partial A_3}{\partial u^1} + \xi_2\frac{\partial A_3}{\partial u^2}) = \xi_1\frac{\partial \boldsymbol{A}}{\partial u^1} + \xi_2\frac{\partial \boldsymbol{A}}{\partial u^2}$$

が成り立つ．結局，$D_{\boldsymbol{X}}\boldsymbol{A}$ は $\boldsymbol{A}$ をスカラーと考えて $\boldsymbol{X}\boldsymbol{A}$ を計算しても同じことになる．

2.1.3　共変微分

ここで，共変微分を導入する．E^3 に埋め込まれた 2 次元曲面 S を引き続き考える．

S 上の点 P における接平面（接空間）を考える．接空間内の任意のベクトル $\boldsymbol{A}$ を考え，この接空間における線形独立な接ベクトル $\boldsymbol{p}_1$ と $\boldsymbol{p}_2$ を用いて，

$$\boldsymbol{A} = A^1\boldsymbol{p}_1 + A^2\boldsymbol{p}_2$$

と表せる．したがって，接ベクトル $\boldsymbol{A}$ の u^1 軸方向への微分 $D_{\frac{\partial}{\partial u^1}}\boldsymbol{A}$ は，

$$\begin{aligned} D_{\frac{\partial}{\partial u^1}}\boldsymbol{A} = \frac{\partial}{\partial u^1}\left(A^1\boldsymbol{p}_1 + A^2\boldsymbol{p}_2\right) &= \frac{\partial A^1}{\partial u^1}\boldsymbol{p}_1 + A_1\frac{\partial \boldsymbol{p}_1}{\partial u^1} + \frac{\partial A^2}{\partial u^1}\boldsymbol{p}_2 + A^2\frac{\partial \boldsymbol{p}_2}{\partial u^1} \\ &= \frac{\partial A^1}{\partial u^1}\boldsymbol{p}_1 + A_1\boldsymbol{p}_{11} + \frac{\partial A^2}{\partial u^1}\boldsymbol{p}_2 + A^2\boldsymbol{p}_{21} \end{aligned} \tag{2.3}$$

である．ここで，

$$\boldsymbol{p}_{11} = \frac{\partial \boldsymbol{p}_1}{\partial u^1} \qquad \boldsymbol{p}_{21} = \frac{\partial \boldsymbol{p}_2}{\partial u^1}$$

の書き方を用いた．

$\boldsymbol{p}_1, \boldsymbol{p}_2$ とこの接平面に対する法ベクトルを $\boldsymbol{n}$ とすれば，$\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{n}$ は線形独立であり，E^3 における基底を構成する．したがって，$\boldsymbol{p}_{ij}$ は $\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{n}$ を基底と考えてこれらの線形結合で

$$\boldsymbol{p}_{ij} = a_{ij}\boldsymbol{p}_1 + b_{ij}\boldsymbol{p}_2 + H_{ij}\boldsymbol{n}$$

と表すことができる．ここで，接ベクトルに対する展開係数 a_{ij}, b_{ij} を，特別な記号 ${\Gamma^k}_{ij}$ を用いて，

$$a_{ij} = {\Gamma^1}_{ij}\,,\ b_{ij} = {\Gamma^2}_{ij}$$

と表す．すると，

$$\boldsymbol{p}_{ij} = \sum_{k=1}^{2}{\Gamma^k}_{ij}\,\boldsymbol{p}_k + H_{ij}\boldsymbol{n} = {\Gamma^k}_{ij}\,\boldsymbol{p}_k + H_{ij}\boldsymbol{n}$$

と置くことができる．${\Gamma^k}_{ij}$ と H_{ij} はインデックス i, j, k に依存した展開係数であるが，S 上の各点 P で与えられており，$\boldsymbol{A}$ を S 上のベクトル場と考えれば，P に依存した関数（C^∞ 級関数）である．すると，$\boldsymbol{A}$ の u^1 方向への微分は，

$$\begin{aligned} D_{\frac{\partial}{\partial u^1}}\boldsymbol{A} &= \frac{\partial A^1}{\partial u^1}\boldsymbol{p}_1 + A^1\left({\Gamma^\mu}_{11}\,\boldsymbol{p}_\mu + H_{11}\boldsymbol{n}\right) + \frac{\partial A^2}{\partial u^1}\boldsymbol{p}_2 + A^2\left({\Gamma^\mu}_{21}\,\boldsymbol{p}_\mu + H_{21}\boldsymbol{n}\right) \\ &= \left(\frac{\partial A^1}{\partial u^1} + A^1{\Gamma^1}_{11} + A^2{\Gamma^1}_{21}\right)\boldsymbol{p}_1 + \left(\frac{\partial A^2}{\partial u^1} + A^1{\Gamma^2}_{11} + A^2{\Gamma^2}_{21}\right)\boldsymbol{p}_2 + \left(A^1H_{11} + A^2H_{21}\right)\boldsymbol{n} \\ &= \left(\frac{\partial A^1}{\partial u^1} + A^\mu{\Gamma^1}_{\mu 1}\right)\boldsymbol{p}_1 + \left(\frac{\partial A^2}{\partial u^1} + A^\mu{\Gamma^2}_{\mu 1}\right)\boldsymbol{p}_2 + \left(A^1H_{11} + A^2H_{21}\right)\boldsymbol{n} \\ &= \left(\frac{\partial A^\nu}{\partial u^1} + A^\mu{\Gamma^\nu}_{\mu 1}\right)\boldsymbol{p}_\nu + A^\mu H_{\mu 1}\boldsymbol{n} \end{aligned}$$

であり，$\boldsymbol{A}$ の u^α 方向への微分は，

$$D_{\frac{\partial}{\partial u^\alpha}}\boldsymbol{A} = \left(\frac{\partial A^\nu}{\partial u^\alpha} + A^\mu{\Gamma^\nu}_{\mu\alpha}\right)\boldsymbol{p}_\nu + A^\mu H_{\mu\alpha}\boldsymbol{n}$$

である．したがって，接成分だけを取れば，

$$D_{\frac{\partial}{\partial u^\alpha}}\boldsymbol{A}\Big|_{\text{接成分}} = \left(\frac{\partial A^\nu}{\partial u^\alpha} + A^\mu \Gamma^\nu{}_{\mu\alpha}\right)\boldsymbol{p}_\nu$$

である．ここで，ベクトル $\boldsymbol{A}$ の方向微分の接成分

$$D_{\frac{\partial}{\partial u^\alpha}}\boldsymbol{A}\Big|_{\text{接成分}}$$

をベクトル $\boldsymbol{A}$ の u^α 方向への共変微分と呼び，記号 $\nabla_{\frac{\partial}{\partial u^\alpha}}\boldsymbol{A}$ で表す．すなわち，

$$\nabla_{\frac{\partial}{\partial u^\alpha}}\boldsymbol{A} = \left(\frac{\partial A^\nu}{\partial u^\alpha} + A^\mu \Gamma^\nu{}_{\mu\alpha}\right)\boldsymbol{p}_\nu \tag{2.4}$$

である．ちなみに，和記号を省略せずに使えば，

$$\nabla_{\frac{\partial}{\partial u^\alpha}}\boldsymbol{A} = \sum_{\nu=1}^{2}\left(\frac{\partial A^\nu}{\partial u^\alpha} + \sum_{\mu=1}^{2} A^\mu \Gamma^\nu{}_{\mu\alpha}\right)\boldsymbol{p}_\nu$$

である．

共変微分 $\nabla_{\boldsymbol{X}}\boldsymbol{Y} = \boldsymbol{Z}$ は，ベクトル場 $\boldsymbol{Y}$ をベクトル場 $\boldsymbol{X}$ の方向への方向微分の接成分であり，その結果 $\boldsymbol{Z}$ もベクトル場である．したがって，共変微分とは 2 つのベクトル場を入力し，1 つのベクトル場を返す写像である．すなわち，曲面 M 上の全てのベクトル場の集合を $\mathfrak{X}(M)$ で表すと，

$$\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \quad \to \quad \mathfrak{X}(M)$$

である．

2.1.4　共変微分の満たす性質

$\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ は曲面上のベクトル場，g, h は曲面上で定義された微分可能なスカラー関数とする．共変微分 $\nabla_{\boldsymbol{X}}\boldsymbol{Y}$ は以下の性質を満たす．

1. $\nabla_{\boldsymbol{X}+\boldsymbol{Y}}\boldsymbol{Z} = \nabla_{\boldsymbol{X}}\boldsymbol{Z} + \nabla_{\boldsymbol{Y}}\boldsymbol{Z}$
2. $\nabla_{g\boldsymbol{X}}\boldsymbol{Z} = g\nabla_{\boldsymbol{X}}\boldsymbol{Z}$
3. $\nabla_{\boldsymbol{X}}(\boldsymbol{Y}+\boldsymbol{Z}) = \nabla_{\boldsymbol{X}}\boldsymbol{Y} + \nabla_{\boldsymbol{X}}\boldsymbol{Z}$
4. $\nabla_{\boldsymbol{X}}(g\boldsymbol{Y}) = (\boldsymbol{X}g)\nabla_{\boldsymbol{X}}\boldsymbol{Y} + g\nabla_{\boldsymbol{X}}\boldsymbol{Y}$

3) の証明

E^3 に埋め込まれた 2 次元の曲面を仮定する．$\boldsymbol{Y} = (Y^1, Y^2, Y^3)$，$\boldsymbol{Z} = (Z^1, Z^2, Z^3)$ とすれば，定義から，

$$\begin{aligned}
D_{\boldsymbol{X}}(\boldsymbol{Y}+\boldsymbol{Z}) &= (\boldsymbol{X}(Y^1+Z^1), \boldsymbol{X}(Y^2+Z^2), \boldsymbol{X}(Y^3+Z^3)) \\
&= (\boldsymbol{X}Y^1+\boldsymbol{X}Z^1, \boldsymbol{X}Y^2+\boldsymbol{X}Z^2, \boldsymbol{X}Y^3+\boldsymbol{X}Z^3) = (\boldsymbol{X}Y^1, \boldsymbol{X}Y^2, \boldsymbol{X}Y^3) + (\boldsymbol{X}Z^1, \boldsymbol{X}Z^2, \boldsymbol{X}Z^3) \\
&= D_{\boldsymbol{X}}\boldsymbol{Y} + D_{\boldsymbol{X}}\boldsymbol{Z}
\end{aligned}$$

である．したがって，両辺の接成分を取れば

$$\nabla_{\boldsymbol{X}}(\boldsymbol{Y}+\boldsymbol{Z}) = \nabla_{\boldsymbol{X}}\boldsymbol{Y} + \nabla_{\boldsymbol{X}}\boldsymbol{Z}$$

が成り立つ．

4) の証明

まず，E^3 における普通の微分 $D_{\boldsymbol{X}}(g\boldsymbol{Y})$ は，

$$D_{\boldsymbol{X}}(g\boldsymbol{Y}) = (\boldsymbol{X}g)\boldsymbol{Y} + gD_{\boldsymbol{X}}\boldsymbol{Y}$$

である．右辺第 1 項である $(\boldsymbol{X}g)\boldsymbol{Y}$ は，接空間のベクトル場であるので，両辺の接成分を考えると

$$D_{\boldsymbol{X}}(g\boldsymbol{Y})\Big|_{\text{接成分}} = (\boldsymbol{X}g)\boldsymbol{Y}\Big|_{\text{接成分}} + gD_{\boldsymbol{X}}\boldsymbol{Y}\Big|_{\text{接成分}} = (\boldsymbol{X}g)\boldsymbol{Y} + g\nabla_{\boldsymbol{X}}\boldsymbol{Y}$$

が成り立つ．また，左辺は

$$D_{\boldsymbol{X}}(g\boldsymbol{Y})\Big|_{\text{接成分}} = \nabla_{\boldsymbol{X}}(g\boldsymbol{Y})$$

であるので，性質４）が証明された．なお，性質１）と２）の証明は省略した．

2.1.5　カッコ積の導入

ここで，カッコ積を導入する．カッコ積は，ベクトル場 $\boldsymbol{X}, \boldsymbol{Y} \in \mathfrak{X}(M)$ が与えられたとき，$f \in C^\infty(M)$ に対して，

$$[\boldsymbol{X}, \boldsymbol{Y}]f = \boldsymbol{X}(\boldsymbol{Y}f) - \boldsymbol{Y}(\boldsymbol{X}f)$$

と作用する微分作用素 $[\boldsymbol{X}, \boldsymbol{Y}]$ をカッコ積という．

ここで，局所座標系を用いて[3]，

$$\boldsymbol{X} = v^i \frac{\partial}{\partial u^i} \qquad \boldsymbol{Y} = w^j \frac{\partial}{\partial u^j}$$

とおけば，

$$\begin{aligned}
\boldsymbol{X}(\boldsymbol{Y}f) &= \left(v^i \frac{\partial}{\partial u^i}\right)\left(w^j \frac{\partial f}{\partial u^j}\right) = v^i \left(\frac{\partial w^j}{\partial u^i}\frac{\partial f}{\partial u^j} + w^j \frac{\partial^2 f}{\partial u^i \partial u^j}\right) \\
\boldsymbol{Y}(\boldsymbol{X}f) &= \left(w^j \frac{\partial}{\partial u^j}\right)\left(v^i \frac{\partial f}{\partial u^i}\right) = w^j \left(\frac{\partial v^i}{\partial u^j}\frac{\partial f}{\partial u^i} + v^i \frac{\partial^2 f}{\partial u^j \partial u^i}\right)
\end{aligned}$$

であるので，

$$\begin{aligned}
[\boldsymbol{X}, \boldsymbol{Y}]f = \boldsymbol{X}(\boldsymbol{Y}f) - \boldsymbol{Y}(\boldsymbol{X}f) &= v^i \frac{\partial w^j}{\partial u^i}\frac{\partial f}{\partial u^j} - w^j \frac{\partial v^i}{\partial u^j}\frac{\partial f}{\partial u^i} \\
&= v^k \frac{\partial w^i}{\partial u^k}\frac{\partial f}{\partial u^i} - w^j \frac{\partial v^i}{\partial u^j}\frac{\partial f}{\partial u^i} = \left(v^k \frac{\partial w^i}{\partial u^k} - w^j \frac{\partial v^i}{\partial u^j}\right)\frac{\partial f}{\partial u^i}
\end{aligned}$$

f は任意の C^∞ 級関数であるので[4]，

$$[\boldsymbol{X}, \boldsymbol{Y}] = \left(v^i \frac{\partial w^k}{\partial u^i} - w^j \frac{\partial v^k}{\partial u^j}\right)\frac{\partial}{\partial u^k}$$

[3] 座標を用いて表示することをこのように言う．

[4] f が C^∞ 級関数であれば偏微分は順序によらないことが示される．すなわち，$\dfrac{\partial^2 f}{\partial u^i \partial u^j} = \dfrac{\partial^2 f}{\partial u^j \partial u^i}$ が成り立つ．

となる．上式は，カッコ積も基底 $\dfrac{\partial}{\partial u^k}$ の線形結合で表されるため，カッコ積も接ベクトル場であることが示される．

次に，

$$[\boldsymbol{X}, \boldsymbol{Y}] = D_{\boldsymbol{X}}\boldsymbol{Y} - D_{\boldsymbol{Y}}\boldsymbol{X}$$

を示す．証明は，

$$D_{\boldsymbol{X}}\boldsymbol{Y}f = D_{\boldsymbol{X}}(\boldsymbol{Y}f) = \boldsymbol{X}(\boldsymbol{Y}f)$$

であるので，

$$D_{\boldsymbol{X}}\boldsymbol{Y}f - D_{\boldsymbol{Y}}\boldsymbol{X}f = \boldsymbol{X}(\boldsymbol{Y}f) - \boldsymbol{Y}(\boldsymbol{X}f) = [\boldsymbol{X}, \boldsymbol{Y}]f$$

となるので，

$$[\boldsymbol{X}, \boldsymbol{Y}] = D_{\boldsymbol{X}}\boldsymbol{Y} - D_{\boldsymbol{Y}}\boldsymbol{X}$$

が成り立つ．さらに，上式の両辺の接空間成分を取れば

$$D_{\boldsymbol{X}}\boldsymbol{Y}\Big|_{\text{接成分}} - D_{\boldsymbol{Y}}\boldsymbol{X}\Big|_{\text{接成分}} = [\boldsymbol{X}, \boldsymbol{Y}]\Big|_{\text{接成分}}$$

であるが，カッコ積は接ベクトル場であるので，結局，

$$\nabla_{\boldsymbol{X}}\boldsymbol{Y} - \nabla_{\boldsymbol{Y}}\boldsymbol{X} = [\boldsymbol{X}, \boldsymbol{Y}]$$

が成り立つ．

2.1.6　ベクトルの内積に対する性質

E^3 に埋め込まれた 2 次元曲面 M 上のベクトル場 $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ を考える．E^3 においては自然な内積が定義でき，

$$\boldsymbol{Y} = (Y^1, Y^2, Y^3) \qquad \boldsymbol{Z} = (Z^1, Z^2, Z^3)$$

とすれば，内積は

$$(\boldsymbol{Y}, \boldsymbol{Z}) = Y^1Z^1 + Y^2Z^2 + Y^3Z^3$$

と定義される．したがって，

$$\begin{aligned} D_{\boldsymbol{X}}(\boldsymbol{Y}, \boldsymbol{Z}) &= \boldsymbol{X}(Y^1Z^1+Y^2Z^2+Y^3Z^3) = (\boldsymbol{X}Y^1)Z^1+Y^1(\boldsymbol{X}Z^1)+(\boldsymbol{X}Y^2)Z^2+Y^2(\boldsymbol{X}Z^2)+(\boldsymbol{X}Y^3)Z^3+Y^3(\boldsymbol{X}Z^3) \\ &= (\boldsymbol{X}Y^1)Z^1 + (\boldsymbol{X}Y^2)Z^2 + (\boldsymbol{X}Y^3)Z^3 + Y^1(\boldsymbol{X}Z^1) + Y^2(\boldsymbol{X}Z^2) + Y^3(\boldsymbol{X}Z^3) \\ &= (D_{\boldsymbol{X}}\boldsymbol{Y}, \boldsymbol{Z}) + (\boldsymbol{Y}, D_{\boldsymbol{X}}\boldsymbol{Z}) \end{aligned}$$

を得る．両辺の接空間成分を取れば，

$$D_{\boldsymbol{X}}(\boldsymbol{Y}, \boldsymbol{Z})\Big|_{\text{接成分}} = (D_{\boldsymbol{X}}\boldsymbol{Y}, \boldsymbol{Z})\Big|_{\text{接成分}} + (\boldsymbol{Y}, D_{\boldsymbol{X}}\boldsymbol{Z})\Big|_{\text{接成分}} = (D_{\boldsymbol{X}}\boldsymbol{Y}\Big|_{\text{接成分}}, \boldsymbol{Z}) + (\boldsymbol{Y}, D_{\boldsymbol{X}}\boldsymbol{Z}\Big|_{\text{接成分}})$$

であり，共変微分に関しても（いわゆるライプニッツ則）

$$\nabla_{\boldsymbol{X}}(\boldsymbol{Y}, \boldsymbol{Z}) = (\nabla_{\boldsymbol{X}}\boldsymbol{Y}, \boldsymbol{Z}) + (\boldsymbol{Y}, \nabla_{\boldsymbol{X}}\boldsymbol{Z}) \tag{2.5}$$

が成り立つ．

2.2 多様体におけるベクトル場の微分とリーマン接続

2.2.1 共変微分の導入

抽象的な曲面では，曲面が高次のユークリッド空間中に存在するとの仮定を置いた議論はできず，内在幾何学に基づいた議論を行う必要がある．

前節では，E^3 に埋め込まれた 2 次元曲面に対して，まず，曲面上でベクトル場の方向微分を定義し，方向微分の接空間内の成分を共変微分として定義した．そして，共変微分が満たす性質について調べた．本節では，この前節での議論をひっくり返し，前節で述べた共変微分の性質を満たす写像として多様体 M 上の共変微分を定義する．

定義 抽象的な曲面（多様体）M 上において，以下の 4 つの条件を満たす写像

$$\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M) \quad : \quad (X, Y) \mapsto \nabla_{\boldsymbol{X}}$$

を M 上の共変微分と呼ぶ．

M 上のベクトル場を $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \in \mathfrak{X}(M)$ とし，任意の微分可能な（C^∞ 級）関数を f とする．

(i) $\nabla_{\boldsymbol{X}+\boldsymbol{Y}}\boldsymbol{Z} = \nabla_{\boldsymbol{X}}\boldsymbol{Z} + \nabla_{\boldsymbol{Y}}\boldsymbol{Z}$

(ii) $\nabla_{f\boldsymbol{X}}\boldsymbol{Z} = f\nabla_{\boldsymbol{X}}\boldsymbol{Z}$

(iii) $\nabla_{\boldsymbol{X}}(\boldsymbol{Y}+\boldsymbol{Z}) = \nabla_{\boldsymbol{X}}\boldsymbol{Y} + \nabla_{\boldsymbol{X}}\boldsymbol{Z}$

(iv) $\nabla_{\boldsymbol{X}}(f\boldsymbol{Y}) = (\boldsymbol{X}f)\nabla_{\boldsymbol{X}}\boldsymbol{Y} + f\nabla_{\boldsymbol{X}}\boldsymbol{Y}$

ここで，ベクトル場 $\boldsymbol{X}$ と $\boldsymbol{Y}$ に対して，写像

$$(X, Y) \quad \mapsto \quad \nabla_{\boldsymbol{X}}$$

が上記 (i) から (iv) を満たせば,

$$\nabla_{\frac{\partial}{\partial u^i}} \frac{\partial}{\partial u^j} = \Gamma^k{}_{ij} \frac{\partial}{\partial u^k} \tag{2.6}$$

なる $\Gamma^k{}_{ij}$ を定めれば $\nabla_{\boldsymbol{X}}\boldsymbol{Y}$ が計算できる．

実際，M 上のベクトル場の基底は

$$\frac{\partial}{\partial u^i} \quad i = 1, \ldots, n$$

で表される．したがって，M 上のベクトル場 $\boldsymbol{X}, \boldsymbol{Y}$ を

$$\boldsymbol{X} = X^i \frac{\partial}{\partial u^i} \qquad \boldsymbol{Y} = Y^j \frac{\partial}{\partial u^j}$$

と表すことができる．すると，上に述べた性質 (i)–(iv) を用いれば,

$$\nabla_{\boldsymbol{X}}\boldsymbol{Y} = \nabla_{X^i \frac{\partial}{\partial u^i}}\boldsymbol{Y} = X^i \nabla_{\frac{\partial}{\partial u^i}}\boldsymbol{Y} = X^i \nabla_{\frac{\partial}{\partial u^i}} \left(Y^j \frac{\partial}{\partial u^j} \right) = X^i \left(\frac{\partial Y^j}{\partial u^i} \frac{\partial}{\partial u^j} + Y^j \nabla_{\frac{\partial}{\partial u^i}} \frac{\partial}{\partial u^j} \right)$$

となる．ここで，式 (2.6) にしたがって $\Gamma^k{}_{ij}$ が定められているとすれば,

$$\begin{aligned}
\nabla_{\boldsymbol{X}}\boldsymbol{Y} = X^i \left(\frac{\partial Y^j}{\partial u^i} \frac{\partial}{\partial u^j} + Y^j \nabla_{\frac{\partial}{\partial u^i}} \frac{\partial}{\partial u^j} \right) &= X^i \left(\frac{\partial Y^j}{\partial u^i} \frac{\partial}{\partial u^j} + Y^j \Gamma^k{}_{ij} \frac{\partial}{\partial u^k} \right) \\
&= \left(X^i \frac{\partial Y^k}{\partial u^i} + X^i Y^j \Gamma^k{}_{ij} \right) \frac{\partial}{\partial u^k}
\end{aligned}$$

を得る．

多様体 M の局所座標系に式 (2.6) を満たす微分可能な（C^∞ 級）関数の組 ${\Gamma^k}_{ij}$ を与えることを，M にアフィン接続を与えると言い，${\Gamma^k}_{ij}$ を接続係数とよぶ．アフィン接続を与えることと，共変微分（すなわち上の定義による写像）を与えることは等価であり，通常,「共変微分」と「アフィン接続」は同じ意味を持つと考える（同一視する）．

2.2.2　捩率に対する条件の追加

多様体 M における共変微分の満たすべき条件に

$$\nabla_{\boldsymbol{X}}\boldsymbol{Y} - \nabla_{\boldsymbol{Y}}\boldsymbol{X} = [\boldsymbol{X}, \boldsymbol{Y}] \tag{2.7}$$

を追加する．実際,

$$\nabla_{\boldsymbol{X}}\boldsymbol{Y} = \left(X^i \frac{\partial Y^k}{\partial u^i} + X^i Y^j {\Gamma^k}_{ij} \right) \frac{\partial}{\partial u^k}$$

$$\nabla_{\boldsymbol{Y}}\boldsymbol{X} = \left(Y^j \frac{\partial Y^k}{\partial u^j} + X^i Y^j {\Gamma^k}_{ji} \right) \frac{\partial}{\partial u^k}$$

であるので,

$$\nabla_{\boldsymbol{X}}\boldsymbol{Y} - \nabla_{\boldsymbol{Y}}\boldsymbol{X} = \left(X^i \frac{\partial Y^k}{\partial u^i} - Y^j \frac{\partial Y^k}{\partial u^j} \right) \frac{\partial}{\partial u^k} + X^i Y^j \left({\Gamma^k}_{ij} - {\Gamma^k}_{ji} \right) \frac{\partial}{\partial u^k}$$

ここで,

$$\left(X^i \frac{\partial Y^k}{\partial u^i} - Y^j \frac{\partial Y^k}{\partial u^j} \right) \frac{\partial}{\partial u^k} = [\boldsymbol{X}, \boldsymbol{Y}]$$

であるので，結局,

$$\nabla_{\boldsymbol{X}}\boldsymbol{Y} - \nabla_{\boldsymbol{Y}}\boldsymbol{X} = [\boldsymbol{X}, \boldsymbol{Y}] + X^i Y^j \left({\Gamma^k}_{ij} - {\Gamma^k}_{ji} \right) \frac{\partial}{\partial u^k}$$

である．式 (2.7) が成り立つ条件から，アフィン接続係数が

$${\Gamma^k}_{ij} = {\Gamma^k}_{ji}$$

を満たさなければならない．

補足

$$T(\boldsymbol{X}, \boldsymbol{Y}) = \nabla_{\boldsymbol{X}}\boldsymbol{Y} - \nabla_{\boldsymbol{Y}}\boldsymbol{X} - [\boldsymbol{X}, \boldsymbol{Y}]$$

はベクトル場を 2 個取って 1 個のベクトル場を返す写像,

$$T : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$$

すなわち，$(0, 2)$ 型のテンソルであり，捩率と呼ばれる．式 (2.7) は，捩率がゼロ,

$$T(\boldsymbol{X}, \boldsymbol{Y}) = 0$$

に対応する．

2.2.3 ベクトルの内積に対する要請

多様体 M 上に計量（接空間に内積）が定義されている場合には，内積に関する要請（式 (2.5)）を満たすとすれば，

$$\nabla_{\boldsymbol{X}}(\boldsymbol{Y},\boldsymbol{Z}) = (\nabla_{\boldsymbol{X}}\boldsymbol{Y},\boldsymbol{Z}) + (\boldsymbol{Y},\nabla_{\boldsymbol{X}}\boldsymbol{Z}) \tag{2.8}$$

$$\nabla_{\boldsymbol{Y}}(\boldsymbol{Z},\boldsymbol{X}) = (\nabla_{\boldsymbol{Y}}\boldsymbol{Z},\boldsymbol{X}) + (\boldsymbol{Z},\nabla_{\boldsymbol{Y}}\boldsymbol{X}) \tag{2.9}$$

$$\nabla_{\boldsymbol{Z}}(\boldsymbol{X},\boldsymbol{Y}) = (\nabla_{\boldsymbol{Z}}\boldsymbol{X},\boldsymbol{Y}) + (\boldsymbol{X},\nabla_{\boldsymbol{Z}}\boldsymbol{Y}) \tag{2.10}$$

が成り立つ．ここで，式 (2.8) と (2.9) を加え，(2.10) を引けば，

$$\nabla_{\boldsymbol{X}}(\boldsymbol{Y},\boldsymbol{Z})+\nabla_{\boldsymbol{Y}}(\boldsymbol{Z},\boldsymbol{X})-\nabla_{\boldsymbol{Z}}(\boldsymbol{X},\boldsymbol{Y}) = (\nabla_{\boldsymbol{X}}\boldsymbol{Y},\boldsymbol{Z})+(\boldsymbol{Y},\nabla_{\boldsymbol{X}}\boldsymbol{Z})+(\nabla_{\boldsymbol{Y}}\boldsymbol{Z},\boldsymbol{X})+(\boldsymbol{Z},\nabla_{\boldsymbol{Y}}\boldsymbol{X})-(\nabla_{\boldsymbol{Z}}\boldsymbol{X},\boldsymbol{Y})-(\boldsymbol{X},\nabla_{\boldsymbol{Z}}\boldsymbol{Y})$$

である．上式の右辺は

$$\text{右辺} = (\nabla_{\boldsymbol{X}}\boldsymbol{Y}+\nabla_{\boldsymbol{Y}}\boldsymbol{X},\boldsymbol{Z}) + (\nabla_{\boldsymbol{X}}\boldsymbol{Z}-\nabla_{\boldsymbol{Z}}\boldsymbol{X},\boldsymbol{Y}) + (\nabla_{\boldsymbol{Y}}\boldsymbol{Z}-\nabla_{\boldsymbol{Z}}\boldsymbol{Y},\boldsymbol{X})$$

となる．ここで，

$$\nabla_{\boldsymbol{X}}\boldsymbol{Y}-\nabla_{\boldsymbol{Y}}\boldsymbol{X} = [\boldsymbol{X},\boldsymbol{Y}]$$

を用いれば，右辺の第 2 項，第 3 項はカッコ積を用いて

$$\begin{aligned}\text{右辺} &= (\nabla_{\boldsymbol{X}}\boldsymbol{Y}+\nabla_{\boldsymbol{Y}}\boldsymbol{X},\boldsymbol{Z}) + ([\boldsymbol{X},\boldsymbol{Z}],\boldsymbol{Y}) + ([\boldsymbol{Y},\boldsymbol{Z}],\boldsymbol{X})\\ &= 2(\nabla_{\boldsymbol{X}}\boldsymbol{Y},\boldsymbol{Z}) - (\nabla_{\boldsymbol{X}}\boldsymbol{Y}-\nabla_{\boldsymbol{Y}}\boldsymbol{X},\boldsymbol{Z}) + ([\boldsymbol{X},\boldsymbol{Z}],\boldsymbol{Y}) + ([\boldsymbol{Y},\boldsymbol{Z}],\boldsymbol{X})\\ &= 2(\nabla_{\boldsymbol{X}}\boldsymbol{Y},\boldsymbol{Z}) - ([\boldsymbol{X},\boldsymbol{Y}],\boldsymbol{Z}) + ([\boldsymbol{X},\boldsymbol{Z}],\boldsymbol{Y}) + ([\boldsymbol{Y},\boldsymbol{Z}],\boldsymbol{X})\end{aligned}$$

である．したがって，

$$2(\nabla_{\boldsymbol{X}}\boldsymbol{Y},\boldsymbol{Z}) = \nabla_{\boldsymbol{X}}(\boldsymbol{Y},\boldsymbol{Z}) + \nabla_{\boldsymbol{Y}}(\boldsymbol{Z},\boldsymbol{X}) - \nabla_{\boldsymbol{Z}}(\boldsymbol{X},\boldsymbol{Y}) + ([\boldsymbol{X},\boldsymbol{Y}],\boldsymbol{Z}) + ([\boldsymbol{Z},\boldsymbol{X}],\boldsymbol{Y}) - ([\boldsymbol{Y},\boldsymbol{Z}],\boldsymbol{X}) \tag{2.11}$$

を得る．

上式は，共変微分 $\nabla_{\boldsymbol{X}}\boldsymbol{Y}$ と任意のベクトル場 $\boldsymbol{Z}$ との内積が右辺のように計算できる，すなわち，接続が確定することを意味している．実際，接続係数は以下のように計算できる．

$$\boldsymbol{X} = \frac{\partial}{\partial u^i} \quad \boldsymbol{Y} = \frac{\partial}{\partial u^j} \quad \boldsymbol{Z} = \frac{\partial}{\partial u^k}$$

として，式 (2.11) に代入する．まず，

$$(\nabla_{\boldsymbol{X}}\boldsymbol{Y},\boldsymbol{Z}) = (\nabla_{\frac{\partial}{\partial u^i}}\frac{\partial}{\partial u^j},\frac{\partial}{\partial u^k}) = ({\Gamma^m}_{ij}\frac{\partial}{\partial u^m},\frac{\partial}{\partial u^k}) = {\Gamma^m}_{ij}\,(\frac{\partial}{\partial u^m},\frac{\partial}{\partial u^k}) = {\Gamma^m}_{ij}\,g_{mk}$$

であり，

$$\nabla_{\boldsymbol{X}}(\boldsymbol{Y},\boldsymbol{Z}) + \nabla_{\boldsymbol{Y}}(\boldsymbol{Z},\boldsymbol{X}) - \nabla_{\boldsymbol{Z}}(\boldsymbol{X},\boldsymbol{Y}) = \frac{\partial g_{jk}}{\partial u^i} + \frac{\partial g_{ki}}{\partial u^j} - \frac{\partial g_{ij}}{\partial u^k}$$

である．さらに，$[\frac{\partial}{\partial u^i},\frac{\partial}{\partial u^j}] = 0$ であるので，

$${\Gamma^m}_{ij}\,g_{mk} = \frac{\partial g_{jk}}{\partial u^i} + \frac{\partial g_{ki}}{\partial u^j} - \frac{\partial g_{ij}}{\partial u^k}$$

あるいは

$${\Gamma^m}_{ij} = g^{mk}\left(\frac{\partial g_{jk}}{\partial u^i} + \frac{\partial g_{ki}}{\partial u^j} - \frac{\partial g_{ij}}{\partial u^k}\right)$$

を得る．上式は接続係数が 1 つに決まり，上式で与えられることを示している．

2.2.4　まとめ：リーマン接続

多様体 M 上のベクトル場を $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \in \mathfrak{X}(M)$ とし，任意の C^∞ 級関数を f とする．

(i) $\nabla_{\boldsymbol{X}+\boldsymbol{Y}}\boldsymbol{Z} = \nabla_{\boldsymbol{X}}\boldsymbol{Z} + \nabla_{\boldsymbol{Y}}\boldsymbol{Z}$

(ii) $\nabla_{f\boldsymbol{X}}\boldsymbol{Z} = f\nabla_{\boldsymbol{X}}\boldsymbol{Z}$

(iii) $\nabla_{\boldsymbol{X}}(\boldsymbol{Y}+\boldsymbol{Z}) = \nabla_{\boldsymbol{X}}\boldsymbol{Y} + \nabla_{\boldsymbol{X}}\boldsymbol{Z}$

(iv) $\nabla_{\boldsymbol{X}}(f\boldsymbol{Y}) = (\boldsymbol{X}f)\nabla_{\boldsymbol{X}}\boldsymbol{Y} + f\nabla_{\boldsymbol{X}}\boldsymbol{Y}$

を満たす写像

$$\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M) \quad : \quad (X, Y) \mapsto \nabla_{\boldsymbol{X}}$$

を M 上の共変微分と呼ぶ．また，上の 4 条件によって決まる接続

$$\nabla_{u^i}\frac{\partial}{\partial u^j} = \Gamma^{\ell}{}_{ij}\,\frac{\partial}{\partial u^\ell}$$

をアフィン接続とよび，$\Gamma^{\ell}{}_{ij}$ をアフィン接続係数と呼ぶ.

（上記 4 条件のみではアフィン接続係数には任意性があり，1 つには決まらない．証明は？）

多様体 M における共変微分の満たすべき条件に

$$\nabla_{\boldsymbol{X}}\boldsymbol{Y} - \nabla_{\boldsymbol{Y}}\boldsymbol{X} = [\boldsymbol{X}, \boldsymbol{Y}]$$

を追加すると，アフィン接続係数は

$$\Gamma^{k}{}_{ij} = \Gamma^{k}{}_{ji}$$

の性質を持つ．さらに，多様体 M 上に計量（接空間に内積）が定義されている場合には，内積に関する要請

$$\nabla_{\boldsymbol{X}}(\boldsymbol{Y}, \boldsymbol{Z}) = (\nabla_{\boldsymbol{X}}\boldsymbol{Y}, \boldsymbol{Z}) + (\boldsymbol{Y}, \nabla_{\boldsymbol{X}}\boldsymbol{Z})$$

も満たすとすれば，接続係数は 1 つに決まり，

$$\Gamma^{m}{}_{ij} = g^{mk}\left(\frac{\partial g_{jk}}{\partial u^i} + \frac{\partial g_{ki}}{\partial u^j} - \frac{\partial g_{ij}}{\partial u^k}\right) \tag{2.12}$$

で表される．上記 5 つの条件を満たすアフィン接続を（つまり式 (2.12) で表される接続係数を持つ接続を）リーマン接続，あるいは Levi-Civita 接続と呼ぶ.

2.2.5　アフィン接続係数–説明補足

アフィン接続係数とは

リーマン空間は，一般に曲がっている．接空間とは，この曲がった空間の局所的な線形近似である．接空間は各点で異なる．ある点の接空間と，わずかに離れた点の接空間を関連付ける写像が必要である．その写像を接続と呼ぶ.

微小距離離れた 2 点 P と $P' = P + dP$ をとり，それらの座標を $\boldsymbol{\xi}, \boldsymbol{\xi} + d\boldsymbol{\xi}$ とする．2 つの点における接空間 $T_P = T_{\boldsymbol{\xi}}$ と $T_{P'} = T_{\boldsymbol{\xi}+d\boldsymbol{\xi}}$ は微妙に異なっている．座標軸方向の接線ベクトル（基底ベクトル）を

$$\boldsymbol{e}_i(\boldsymbol{\xi}) \in T_P \qquad \boldsymbol{e}'_i = \boldsymbol{e}_(\boldsymbol{\xi} + d\boldsymbol{\xi}) \in T'_P$$

と書く．ここで問題は，$\boldsymbol{e}'$ は $T_{P'}$ のベクトルであって，T_P のベクトルでないから，$\boldsymbol{e}_i$ と $\boldsymbol{e}'_i$ は異なるベクトル空間に属していて直ちにこれらを比較したりは出来ない．

接続とは，$T_{P'}$ のベクトルを T_P のベクトルに写して対応するベクトルを定めることである．$\boldsymbol{e}'_i \in T_{P'}$ を T_P に写すと，T_P では $\widetilde{e}_i$ が対応するとする．すなわち，

$$T_{P'} \ni \boldsymbol{e}'_i \mapsto \widetilde{\boldsymbol{e}}_i \in T_P$$

である．上式で示す対応が線形であるとき，この対応をアフィン接続と呼ぶ．$\widetilde{e}_i$ は $\boldsymbol{e}_i$ に近いであろうから

$$\widetilde{\boldsymbol{e}}_i = \boldsymbol{e}_i + d\boldsymbol{e}_i$$

と書ける．線形な対応を仮定したアフィン接続ではこの $d\boldsymbol{e}_i$ が T_P の基底ベクトル $\boldsymbol{e}_i$ を使って，

$$d\boldsymbol{e}_i = (de_i^k)\boldsymbol{e}_k \quad \left(= \sum_{k=1}^{n} (de_i^k)\boldsymbol{e}_k\right)$$

と書ける．つまり，de_i^k はベクトル $d\boldsymbol{e}_i$ の k 番目の基底に対する成分である．

ここで，微小な線素（ベクトル）$d\boldsymbol{\xi}$ がゼロになれば $\boldsymbol{e}'_i$ は $\boldsymbol{e}_i$ に一致するから，$d\boldsymbol{e}_i$ は線形近似で $\boldsymbol{\xi}$ に比例している．そこで，ある係数 ${\Gamma^k}_{ij}$ を用いて，

$$de_i^k = {\Gamma^k}_{ij}\, d\xi^j \quad \text{あるいは} \quad d\boldsymbol{e}_i = {\Gamma^k}_{ij}\, d\xi^j \boldsymbol{e}_k$$

と書くことができる．すなわち，係数の組 ${\Gamma^k}_{ij}$ が定まっていれば，$T_{P'}$ のベクトルを T_P のベクトルに変換できる．係数の組 ${\Gamma^k}_{ij}$ を接続の係数とよぶ．上の議論では，写像 $T_{P'} \to T_P$ に線形性を仮定している．線形性を仮定して求まった接続をアフィン接続と呼ぶ．

$d\boldsymbol{e}_i$ と $\boldsymbol{e}_m$ の内積を取ると，

$$\langle d\boldsymbol{e}_i, \boldsymbol{e}_m\rangle = \langle {\Gamma^k}_{ij}\, d\xi^j \boldsymbol{e}_k, \boldsymbol{e}_m\rangle = {\Gamma^k}_{ij}\, d\xi^j \langle \boldsymbol{e}_k, \boldsymbol{e}_m\rangle = {\Gamma^k}_{ij}\, d\xi^j g_{km} = \Gamma_{ij,m} d\xi^j$$

を得る．

ベクトル場の方向微分

スカラー関数がベクトル場でどのように変化するかは，スカラー関数の方向微分（軸方向なら単なる偏微分）を計算すればよい．それでは，ベクトル場でベクトルが場所によってどう変わっていくかを計算してみる．

ここでは，

$$\boldsymbol{X}(\boldsymbol{\xi}) = X^i(\boldsymbol{\xi})\boldsymbol{e}_i(\boldsymbol{\xi}) \quad \text{および} \quad \boldsymbol{X}(\boldsymbol{\xi} + d\boldsymbol{\xi}) = X^i(\boldsymbol{\xi} + d\boldsymbol{\xi})\boldsymbol{e}_i(\boldsymbol{\xi} + d\boldsymbol{\xi})$$

の違いを見るわけである．これら 2 つのベクトルは成分だけでなく基底も異なるので，違いを見るには $T_{\boldsymbol{\xi}+d\boldsymbol{\xi}}$ のベクトル $\boldsymbol{X}(\boldsymbol{\xi} + d\boldsymbol{\xi})$ を $T_{\boldsymbol{\xi}}$ に移して違いを見なければならない．$\boldsymbol{e}_i(\boldsymbol{\xi} + d\boldsymbol{\xi}) \in T_{\boldsymbol{\xi}+d\boldsymbol{\xi}}$ を $T_{\boldsymbol{\xi}}$ に移したものは，

$$\widetilde{\boldsymbol{e}}_i(\boldsymbol{\xi} + d\boldsymbol{\xi}) = \boldsymbol{e}_i(\boldsymbol{\xi}) + {\Gamma^k}_{ji}\, d\xi^i \boldsymbol{e}_k(\boldsymbol{\xi})$$

であるので，$\boldsymbol{X}(\boldsymbol{\xi})$ の ξ^i 軸方向の微分（$\boldsymbol{e}_i$ 方向への微分）は，まず，

$$X^i(\boldsymbol{\xi}+d\boldsymbol{\xi})\widetilde{e}_i(\boldsymbol{\xi}+d\boldsymbol{\xi})-X^i(\boldsymbol{\xi})\boldsymbol{e}_i(\boldsymbol{\xi})$$

を計算する．したがって，

$$\begin{aligned}X^i(\boldsymbol{\xi}+d\boldsymbol{\xi})\widetilde{e}_i(\boldsymbol{\xi}+d\boldsymbol{\xi})-X^i(\boldsymbol{\xi})\boldsymbol{e}_i(\boldsymbol{\xi}) &= X^i(\boldsymbol{\xi}+d\boldsymbol{\xi})\left[\boldsymbol{e}_i(\boldsymbol{\xi})+\Gamma^k{}_{ji}\,d\xi^i\boldsymbol{e}_k(\boldsymbol{\xi})\right]-X^i(\boldsymbol{\xi})\boldsymbol{e}_i(\boldsymbol{\xi})\\ &= \left[X^i(\boldsymbol{\xi}+d\boldsymbol{\xi})-X^i(\boldsymbol{\xi})\right]\boldsymbol{e}_i(\boldsymbol{\xi})+X^i(\boldsymbol{\xi}+d\boldsymbol{\xi})\Gamma^k{}_{ji}\,d\xi^j\boldsymbol{e}_k(\boldsymbol{\xi})\end{aligned}$$

である．ここで，ベクトル場 $\boldsymbol{X}=X^k\boldsymbol{e}_k$ の i 軸方向への変化を $\nabla_i(X^k\boldsymbol{e}_k)$ と書けば，

$$\begin{aligned}\nabla_i(X^k\boldsymbol{e}_k) &= \lim_{d\xi^i\to 0}\frac{1}{d\xi^i}\left(X^m(\boldsymbol{\xi}+d\xi^i)\widetilde{\boldsymbol{e}}_m(\boldsymbol{\xi}+d\boldsymbol{\xi})-X^m(\boldsymbol{\xi})\boldsymbol{e}_m(\boldsymbol{\xi})\right)\\ &= \lim_{d\xi^i\to 0}\frac{1}{d\xi^i}\left[X^m(\boldsymbol{\xi}+d\xi^i)-X^m(\boldsymbol{\xi})\right]\boldsymbol{e}_m(\boldsymbol{\xi})+\lim_{d\xi^i\to 0}\frac{1}{d\xi^i}X^m(\boldsymbol{\xi}+d\xi^i)\Gamma^k{}_{jm}\,d\xi^j\boldsymbol{e}_k(\boldsymbol{\xi})\end{aligned}$$

となる．右辺において，第 1 項は

$$\lim_{d\xi^i\to 0}\frac{1}{d\xi^i}\left[X^m(\boldsymbol{\xi}+d\xi^i)-X^m(\boldsymbol{\xi})\right]\boldsymbol{e}_m(\boldsymbol{\xi})=\partial_iX^k\boldsymbol{e}_k(\boldsymbol{\xi})=\frac{\partial X^k(\boldsymbol{\xi})}{\partial\xi^i}\boldsymbol{e}_k(\boldsymbol{\xi})$$

である．第 2 項は，

$$\lim_{d\xi^i\to 0}\frac{1}{d\xi^i}X^m(\boldsymbol{\xi}+d\xi^i)\Gamma^k{}_{jm}\,d\xi^j\boldsymbol{e}_k(\boldsymbol{\xi})=\lim_{d\xi^i\to 0}\frac{1}{d\xi^i}X^m(\boldsymbol{\xi}+d\xi^i)(\Gamma^k{}_{im}\,d\xi^i)\boldsymbol{e}_k(\boldsymbol{\xi})=X^j(\boldsymbol{\xi})(\Gamma^k{}_{ij})\boldsymbol{e}_k(\boldsymbol{\xi})$$

となる．ここで，右辺の $(\Gamma^k{}_{im}\,d\xi^i)$ において i で和はとらない．なぜならば，この計算は $\boldsymbol{\xi}$ の成分のうち i 番目の ξ^i にのみ $d\xi^i$ の変位を与えたのであるから，

$$\Gamma^k{}_{jm}\,d\xi^j=\sum_{j=1}^{n}\Gamma^k{}_{jm}\,d\xi^j=\Gamma^k{}_{im}\,d\xi^i\quad(i\text{ で和は取らない})$$

となるわけである．なぜなら，

$$d\xi^j=\begin{cases}0 & j\neq i\\ d\xi^i & j=i\end{cases}$$

であるからである．したがって，

$$\nabla_i(X^k\boldsymbol{e}_k)=\left(\partial_iX^k+\Gamma^k{}_{ij}\,X^j\right)\boldsymbol{e}_k$$

を得る．書き方として，

$$\nabla_iX^k=\partial_iX^k+\Gamma^k{}_{ij}\,X^j$$

と書く場合もある．この書き方では，ベクトルの第 k 成分 X^k の i 方向への共変微分が $\partial_iX^k+\Gamma^k{}_{ij}\,X^j$ となるような印象を与えかねないが，あくまで，ベクトル $\boldsymbol{X}$ の i 方向への共変微分（共変微分もベクトルである）を表すベクトルの第 k 成分が $\partial_iX^k+\Gamma^k{}_{ij}\,X^j$ であるとの意味である．

さらに，ベクトル場 $\boldsymbol{X}$ の $\boldsymbol{Y}=Y^i\boldsymbol{e}_i$ 方向への共変微分 $\nabla_Y\boldsymbol{X}$ は

$$\nabla_Y\boldsymbol{X}=\nabla_{Y^i\boldsymbol{e}_i}\boldsymbol{X}=Y^i\nabla_{\boldsymbol{e}_i}\boldsymbol{X}=Y^i\nabla_i\boldsymbol{X}=Y^i\left(\partial_iX^k+\Gamma^k{}_{ij}\,X^j\right)\boldsymbol{e}_k$$

である．基底ベクトル場もベクトル場であるので，

$$\nabla_{\boldsymbol{e}_i}\boldsymbol{e}_j=\Gamma^k{}_{ij}\,\boldsymbol{e}^k$$

や

$$\langle\nabla_{\boldsymbol{e}_i}\boldsymbol{e}_j,\boldsymbol{e}_k\rangle=\Gamma^m{}_{ij}\,\langle\boldsymbol{e}^m,\boldsymbol{e}_k\rangle=\Gamma^m{}_{ij}\,g_{mk}=\Gamma_{ij,k}$$

が成り立つ．

2.3　ベクトルの平行移動

2.3.1　E^3 に埋め込まれた 2 次元曲面の場合

まず，E^3 に埋め込まれた 2 次元曲面の場合を考えてみる．この曲面上の曲線 $C\,:\,\boldsymbol{p}(t)=(x(t),y(t),z(t))$ を仮定する．t は曲線上の位置を表すパラメータである．曲線 C 上に定義されたベクトル場 $\boldsymbol{v}(t)$ を考え，$\boldsymbol{v}(t)$ を C に沿って微分することを考える．つまり，

$$\frac{d\boldsymbol{v}(t)}{dt}=\lim_{\Delta t\to 0}\frac{\boldsymbol{v}(t+\Delta t)-\boldsymbol{v}(t)}{\Delta t}$$

を計算する．ここで上式右辺においては，ベクトル $\boldsymbol{v}(t+\Delta t)$ と $\boldsymbol{v}(t)$ との差分を計算している．これは，曲線 C 上の位置 $\boldsymbol{p}(t+\Delta t)$ にあるベクトル $\boldsymbol{v}(t+\Delta t)$ を，位置 $\boldsymbol{p}(t)$ に平行移動して，その位置にある $\boldsymbol{v}(t)$ と差分を計算していることになる．(図 2.1 参照のこと．)

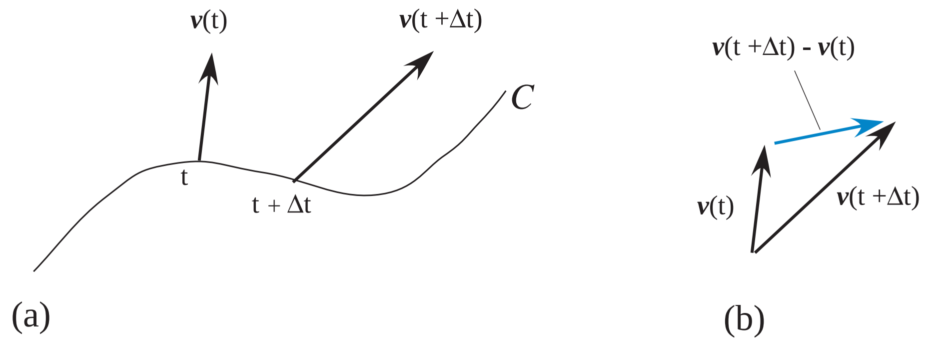

図 2.1:

この「平行移動」はユークリッド空間 E^3 では，あたりまえの操作として無意識に実行している．しかしながら，抽象的な曲面である多様体上では平行移動は「あたりまえ」ではなく，このような形では実行できないのである．多様体 M 上ではベクトル $\boldsymbol{v}(t+\Delta t)$ は接空間 $\mathcal{T}_{\boldsymbol{p}(t+\Delta t)}$ に属しており，ベクトル $\boldsymbol{v}(t)$ は接空間 $\mathcal{T}_{\boldsymbol{p}(t)}$ に属していて，一般的には両者の間に代数的な関係は存在しないので

$$\boldsymbol{v}(t+\Delta t)-\boldsymbol{v}(t)$$

のように直接差分を計算することはできない．(多様体上で位置が異なれば，接空間はそれぞれが異なるベクトル空間であり，それらの基底も異なる．)

もう 1 度 E^3 に埋め込まれた 2 次元曲面の場合に話を戻す．曲線 C 上で定義されたベクトル場を $\boldsymbol{v}(t)=(v^1(t),v^2(t),v^3(t))$ とする．ここで，曲線上の位置 t にかかわらずベクトルが等しい，つまり曲線上のベクトルが平行であれば，a,b,c を定数として，

$$\boldsymbol{v}(t)=\boldsymbol{v}=(a,b,c)$$

となる．この場合，

$$\frac{d\boldsymbol{v}(t)}{dt}=0$$

である．さらに，

$$\frac{d}{dt}=\frac{dx^i}{dt}\frac{\partial}{\partial x^i}$$

であるので，

$$\frac{d\boldsymbol{v}(t)}{dt} = \frac{dx^i}{dt} D_{\frac{\partial}{\partial x^i}} \boldsymbol{v}(t)$$

である．

ここで，$\dfrac{d\boldsymbol{v}(t)}{dt}$ は E^3 のベクトルであるので，C の接線方向の成分と，接線に垂直な方向への成分に分解でき，これら 2 つの方向は線形独立である．したがって，$\dfrac{d\boldsymbol{v}(t)}{dt} = 0$ であれば，当然ながら $\dfrac{d\boldsymbol{v}(t)}{dt}$ の接成分もゼロ，つまり，

$$\frac{d\boldsymbol{v}(t)}{dt}\Big|_{接成分} = \frac{dx^i}{dt} D_{\frac{\partial}{\partial x^i}} \boldsymbol{v}(t) \quad \Big|_{接成分} = \dot{x}^i \nabla_{\frac{\partial}{\partial x^i}} \boldsymbol{v}(t) = 0$$

である．したがって，C 上のベクトル場が平行の場合，

$$\boldsymbol{v}\text{ が一定（平行）ベクトル場} \iff \nabla_{\frac{\partial}{\partial x^i}} \boldsymbol{v} = 0 \iff \frac{\partial v^k}{\partial x^j} + v^i \Gamma^k{}_{ij} = 0 \tag{2.13}$$

が言える．

2.3.2 多様体上でのベクトルの平行移動

E^3 に埋め込まれた 2 次元曲面の場合に，$\boldsymbol{v}$ が一定（平行）ベクトル場であるとき

$$\frac{\partial v^k}{\partial x^j} + v^i \Gamma^k{}_{ij} = 0$$

が成り立つことを上で議論した．多様体上でベクトルの平行移動を議論する場合，この議論を逆にして，ベクトル $\boldsymbol{v}$ が曲線 C に沿って上の式を満たすなら $\boldsymbol{v}$ は曲線 C に沿って一定（平行）ベクトル場であると定義するのである．

ここでの問題は，抽象的な曲面である多様体においては，ユークリッド空間のように等く離れた 2 点でのベクトルを比較できない事である．離れた 2 点間のベクトルが平行かどうかは，2 点を結ぶ経路に依存するためである．すなわち，曲線に沿って接ベクトルを平行移動すると，一般にその結果はその曲線（経路）に依存してしまう．

そこで，まず，ベクトル場 $\boldsymbol{Z}$ がある曲線に沿って平行であることを定義する．

準備として，多様体 M 上に平行移動の経路としてなめらかな曲線 $C = \{\boldsymbol{p}(t)\}$ を仮定する．この曲線は局所座標表示で

$$\boldsymbol{p}(t) = (x^1(t), x^2(t), \ldots, x^n(t))$$

である．$t = t_0$ における曲線 C の速度ベクトル $\dot{\boldsymbol{p}}(t_0) = \left(\dfrac{d\boldsymbol{p}}{dt}\right)_{t_0}$ は

$$\dot{\boldsymbol{p}}(t_0) = \frac{dx^i}{dt}\Big|_{t_0} \left(\frac{\partial}{\partial x^i}\right)_{\boldsymbol{p}(t_0)}$$

である．ここで，接ベクトルと微分作用素を同一視すれば $\dot{\boldsymbol{p}}(t_0)$ は $\left(\dfrac{d}{dt}\right)_{t_0}$ と言う演算子で表される．つまり，

$$\left(\frac{d}{dt}\right)_{t_0} = \frac{dx^i}{dt}\Big|_{t_0} \left(\frac{\partial}{\partial x^i}\right)_{\boldsymbol{p}(t_0)}$$

である．ここで，$t=t_0$ を指定せず，

$$\frac{d}{dt}=\frac{dx^i}{dt}\frac{\partial}{\partial x^i}\quad(=\dot{\boldsymbol{p}}(t))$$

と書けば，接ベクトル場を表す．

定義 多様体 M 上の曲線 $C=\{\boldsymbol{p}(t);a\leq t\leq b\}$ に沿って定義されたベクトル場 $\boldsymbol{Z}$ が

$$\nabla_{\dot{\boldsymbol{p}}(t)}\boldsymbol{Z}=0\qquad a\leq t\leq b \tag{2.14}$$

を満たすとき，ベクトル場 $\boldsymbol{Z}$ は曲線 C に沿って平行であるという．

以上が，ベクトル場の平行性の定義であり，必ずある曲線に対する平行性を考える．つまり，ベクトル場 $\boldsymbol{Z}$ が，ある曲線の接ベクトル（速度ベクトル）の方向に対する共変微分がゼロであるとき，$\boldsymbol{Z}$ はその曲線に対して平行であると言う．

式 (2.14) を局所座標で表示してみる．まず，$\boldsymbol{Z}$ の局所座標表示は，

$$\boldsymbol{Z}=Z^i\frac{\partial}{\partial x^i}$$

であり，

$$\begin{aligned}
\nabla_{\dot{\boldsymbol{p}}(t)}\boldsymbol{Z}=\nabla_{\frac{d}{dt}}\boldsymbol{Z}=\nabla_{\frac{dx^i}{dt}\frac{\partial}{\partial x^i}}\left(Z^j\frac{\partial}{\partial x^j}\right)&=\left(\nabla_{\frac{dx^i}{dt}\frac{\partial}{\partial x^i}}Z^j\right)\frac{\partial}{\partial x^j}+Z^j\left(\nabla_{\frac{dx^i}{dt}\frac{\partial}{\partial x^i}}\frac{\partial}{\partial x^j}\right)\\
&=\frac{dx^i}{dt}\frac{\partial Z^j}{\partial x^i}\frac{\partial}{\partial x^j}+Z^j\frac{dx^i}{dt}\left(\nabla_{\frac{\partial}{\partial x^i}}\frac{\partial}{\partial x^j}\right)=\frac{dZ^j}{dt}\frac{\partial}{\partial x^j}+Z^j\frac{dx^i}{dt}\Gamma^k{}_{ij}\frac{\partial}{\partial x^k}\\
&=\left(\frac{dZ^k}{dt}+Z^j\frac{dx^i}{dt}\Gamma^k{}_{ij}\right)\frac{\partial}{\partial x^k}
\end{aligned}$$

である．したがって，$\nabla_{\dot{\boldsymbol{p}}(t)}\boldsymbol{Z}=0$ 　は

$$\frac{dZ^k}{dt}+Z^j\frac{dx^i}{dt}\Gamma^k{}_{ij}=0\qquad(k=1,\ldots,n) \tag{2.15}$$

と同値である．

式 (2.15) は，$Z^i(t)$ に関する 1 階線形常微分方程式である．したがって，$\boldsymbol{p}(a)\in M$ における接ベクトルを $\boldsymbol{v}$ とすれば，C に沿って平行なベクトル場 $\boldsymbol{Z}_{\boldsymbol{p}(t)}$ で $\boldsymbol{Z}_{\boldsymbol{p}(a)}=\boldsymbol{v}$ であるものが微分方程式 (2.15) の解として求まる．（ベクトル場 $\boldsymbol{Z}$ の C 上の点 $\boldsymbol{p}(t)$ におけるベクトルを $\boldsymbol{Z}_{\boldsymbol{p}(t)}$ と表した．）

M 上の曲線 $C=\{\boldsymbol{p}(t);a\leq t\leq b\}$ と $\boldsymbol{v}\in\mathcal{T}_{\boldsymbol{p}(a)}M$ が与えられたとき，式 (2.15) の解として求まるベクトル場 $\boldsymbol{Z}$ の $t=b$ での値 $\boldsymbol{Z}_{\boldsymbol{p}(b)}\in\mathcal{T}_{\boldsymbol{p}(b)}M$ を，C に沿って $\boldsymbol{v}$ を $\boldsymbol{p}(a)$ から $\boldsymbol{p}(b)$ まで平行移動して得られる接ベクトルと考える．

この平行移動を Π_C と表せば，Π_C は，2 つの接空間 $\mathcal{T}_{\boldsymbol{p}(a)}M$ と $\mathcal{T}_{\boldsymbol{p}(b)}M$ の間の対応関係を定める．すなわち，

$$\Pi_C\,:\,\mathcal{T}_{\boldsymbol{p}(a)}M\quad\rightarrow\quad\mathcal{T}_{\boldsymbol{p}(b)}M$$

である．Π_C は全単射な線形写像である．線形性は微分方程式 (2.15) が Z^i に関して線型方程式であることの帰結であり，全単射であることは微分方程式 (2.15) の解の一意性からの帰結である[5]

[5] このことは証明が必要か．

Π_C を用いると，多様体 M 上で，曲線 $C=\{\boldsymbol{p}(t); a\le t\le b\}$ を介して，離れた 2 点 $\boldsymbol{p}(a)$, $\boldsymbol{p}(b)\in M$ での接ベクトルが平行か否かを論ずることができる．つまり，

$$\boldsymbol{v}\in\mathcal{T}_{\boldsymbol{p}(a)}M \text{ と } \boldsymbol{w}\in\mathcal{T}_{\boldsymbol{p}(b)}M \text{ が } C \text{ を介して平行である．} \quad\Longleftrightarrow\quad \Pi_C\boldsymbol{v}=\boldsymbol{w}$$

が成り立つ．

Π_C を用いると共変微分の幾何学的な意味づけを以下のように与えることができる．

定理 多様体 M において，点 $\boldsymbol{p}(a)\in M$ を通る曲線

$$C=\{\boldsymbol{p}(t) : a-\boldsymbol{\varepsilon}<t<a+\boldsymbol{\varepsilon}\}\qquad \boldsymbol{\varepsilon}>0$$

を考える．任意のベクトル場 $\boldsymbol{Y}\in\mathfrak{X}(M)$ に対して，

$$\left(\nabla_{\dot{\boldsymbol{p}}(t)}\boldsymbol{Y}\right)_{\boldsymbol{p}(a)}=\lim_{t\to a}\frac{1}{t-a}\left(\Pi_{\boldsymbol{p}(a)}^{\boldsymbol{p}(t)}\boldsymbol{Y}_{\boldsymbol{p}(t)}-\boldsymbol{Y}_{\boldsymbol{p}(a)}\right)$$

が成り立つ．($\Pi_{\boldsymbol{p}(a)}^{\boldsymbol{p}(t)}$ は特に始点と終点を明示して書いたものであり，Π_C と同じ意味である.)

証明： 曲線 C を $C=\{\boldsymbol{p}(t)=(x^1(t),\ldots,x^n(t)) : a-\boldsymbol{\varepsilon}\le t\le a+\boldsymbol{\varepsilon}\}$ と局所座標表示して，$\boldsymbol{Y}$ を C 上で成分表示する．

$$\boldsymbol{Y}_{\boldsymbol{p}(t)}=Y^i(t)\left(\frac{\partial}{\partial x^i}\right)_{\boldsymbol{p}(t)}$$

また，$\boldsymbol{Z}$ を C に沿って平行な任意のベクトル場として，

$$\boldsymbol{Z}_{\boldsymbol{p}(t)}=Z^j(t)\left(\frac{\partial}{\partial x^i}\right)_{\boldsymbol{p}(t)}$$

と成分表示する．$\boldsymbol{Z}$ が C に沿って平行であるので式 (2.15)：

$$\frac{dZ^k}{dt}+Z^j\frac{dx^i}{dt}\Gamma^k{}_{ij}=0\qquad (k=1,\ldots,n)$$

が成り立つ．$t=a$ の近傍で上式を少し書き換えて，

$$\frac{Z^k(a+h)-Z^k(a)}{h}+Z^j(a)\dot{x}^i(a)\Gamma^k{}_{ij}\left(\boldsymbol{p}(a)\right)\approx 0 \tag{2.16}$$

が言える．ここで，$\dot{x}^i=dx^i/dt$ である．同程度の近似の範囲内で

$$\frac{Z^k(a+h)-Z^k(a)}{h}+Z^j(a+h)\dot{x}^i(a+h)\Gamma^k{}_{ij}\left(\boldsymbol{p}(a+h)\right)\approx 0 \tag{2.17}$$

も言える．

ここで，C 上の平行なベクトル場 $\boldsymbol{Z}$ を $\boldsymbol{Z}_{\boldsymbol{p}(a+h)}=\boldsymbol{Y}_{\boldsymbol{p}(a+h)}$ とする．すなわち，$\boldsymbol{Z}$ として，ベクトル $\boldsymbol{Y}_{\boldsymbol{p}(a+h)}$ に平行なベクトル場とする．

すると，

$$\Pi_{\boldsymbol{p}(a)}^{\boldsymbol{p}(a+h)}\boldsymbol{Y}_{\boldsymbol{p}(a+h)}=\Pi_{\boldsymbol{p}(a)}^{\boldsymbol{p}(a+h)}\boldsymbol{Z}_{\boldsymbol{p}(a+h)}=\boldsymbol{Z}_{\boldsymbol{p}(a)}=Z^k(a)\left(\frac{\partial}{\partial x^k}\right)_{\boldsymbol{p}(a)}$$

である．ここで，$Z^k(a)$ は，式 (2.17) より，

$$Z^k(a)=Z^k(a+h)+hZ^j(a+h)\dot{x}^i(a+h)\Gamma^k{}_{ij}\left(\boldsymbol{p}(a+h)\right)+O(h)$$

である．ここで，$O(h)$ は有限な h に伴なう近似誤差を表す．ここで，$\boldsymbol{Z}_{\boldsymbol{p}(a+h)} = \boldsymbol{Y}_{\boldsymbol{p}(a+h)}$ であることより，

$$Z^k(a) = Y^k(a+h) + hY^j(a+h)\dot{x}^i(a+h){\Gamma^k}_{ij}\left(\boldsymbol{p}(a+h)\right) + O(h)$$

である．したがって，

$$\Pi_{\boldsymbol{p}(a)}^{\boldsymbol{p}(a+h)}\boldsymbol{Y}_{\boldsymbol{p}(a+h)} = Z^k(a)\left(\frac{\partial}{\partial x^k}\right)_{\boldsymbol{p}(a)} = \left(Y^k(a+h) + hY^j(a+h)\dot{x}^i(a+h){\Gamma^k}_{ij}\left(\boldsymbol{p}(a+h)\right) + O(h)\right)\left(\frac{\partial}{\partial x^k}\right)_{\boldsymbol{p}(a)}$$

であり，

$$\begin{aligned}
&\frac{1}{h}\left(\Pi_{\boldsymbol{p}(a)}^{\boldsymbol{p}(a+h)}\boldsymbol{Y}_{\boldsymbol{p}(a+h)} - \boldsymbol{Y}_{\boldsymbol{p}(a)}\right) \\
&\quad = \frac{1}{h}\left[\left(Y^k(a+h) + hY^j(a+h)\dot{x}^i(a+h){\Gamma^k}_{ij}\left(\boldsymbol{p}(a+h)\right) + O(h)\right)\left(\frac{\partial}{\partial x^k}\right)_{\boldsymbol{p}(a)} - Y^k(a)\left(\frac{\partial}{\partial x^k}\right)_{\boldsymbol{p}(a)}\right] \\
&\qquad = \left(\frac{Y^k(a+h) - Y^k(a)}{h} + Y^j(a+h)\dot{x}^i(a+h){\Gamma^k}_{ij}\left(\boldsymbol{p}(a+h)\right) + O(h)\right)\left(\frac{\partial}{\partial x^k}\right)_{\boldsymbol{p}(a)}
\end{aligned}$$

であるので，

$$\lim_{h\to 0}\frac{1}{h}\left(\Pi_{\boldsymbol{p}(a)}^{\boldsymbol{p}(a+h)}\boldsymbol{Y}_{\boldsymbol{p}(a+h)} - \boldsymbol{Y}_{\boldsymbol{p}(a)}\right) = \left(\frac{dY^k}{dt}(a) + Y^j(a)\dot{x}^i(a){\Gamma^k}_{ij}\left(\boldsymbol{p}(a)\right)\right)\left(\frac{\partial}{\partial x^k}\right)_{\boldsymbol{p}(a)} = \left(\nabla_{\dot{\boldsymbol{p}}(t)}\boldsymbol{Y}\right)_{\boldsymbol{p}(a)}$$

となる．

すなわち，上式左辺が共変微分の意味であり，接空間 $\mathcal{T}_{\boldsymbol{p}(a+h)}$ に属するベクトル $\boldsymbol{Y}_{\boldsymbol{p}(a+h)} \in \mathcal{T}_{\boldsymbol{p}(a+h)}$ を，接空間 $\mathcal{T}_{\boldsymbol{p}(a)}$ へ平行移動して，この接空間の元であるベクトル $\boldsymbol{Y}_{\boldsymbol{p}(a)}$ との間で差分を計算している．これは，ユークリッド空間での（通常の）ベクトルの微分

$$\frac{d\boldsymbol{v}(t)}{dt} = \lim_{\Delta t\to 0}\frac{\boldsymbol{v}(t+\Delta t) - \boldsymbol{v}(t)}{\Delta t}$$

の，多様体という抽象的な曲面への拡張であることがわかる．

2.3.3　補足：測地線

多様体 M 上の曲線 C に沿ったベクトル場 $\boldsymbol{Z}$ を考えた．$\boldsymbol{Z}$ が C に関して平行であるとは式 (2.15)，すなわち，

$$\frac{dZ^k}{dt} + Z^j\frac{dx^i}{dt}{\Gamma^k}_{ij} = 0 \qquad (k = 1, \ldots, n)$$

が成り立つことが条件であった．ここで，曲線 C の速度ベクトル（接ベクトル）$\dot{\boldsymbol{p}}(t)$ について考えてみる．この $\dot{\boldsymbol{p}}(t)$ も C に沿ったベクトル場をなす．ここで，

$$\nabla_{\dot{\boldsymbol{p}}(t)}\dot{\boldsymbol{p}}(t) = 0$$

は，C の速度ベクトル場が C 自身に沿って平行であることを意味する．このような曲線を測地線と呼ぶ．座標表示では，$Z^k = \dot{x}^k$ を式 (2.15) に代入し，

$$\frac{d^2x^k}{dt^2} + {\Gamma^k}_{ij}\frac{dx^i}{dt}\frac{dx^j}{dt} = 0 \qquad (k = 1, \ldots, n)$$

を得る．上式は測地線方程式と呼ばれる．

第3章　曲率から双対アフィン接続まで

多様体の曲率，接続を復習し，情報幾何で用いられる双対アフィン接続について説明する．

3.1　曲率テンソル

3.1.1　曲率テンソルの定義

多様体の上でベクトル場 $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \in \mathfrak{X}(M)$ に対して，$\nabla_X \nabla_Y \boldsymbol{Z}$ と $\nabla_Y \nabla_X \boldsymbol{Z}$ は同じにはならない．すなわち，$\nabla_X \nabla_Y$ と $\nabla_Y \nabla_X$ は交換可能ではない．この差が曲面の曲がり方に関係する．

$$R : \mathfrak{X}(M) \times \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$$

に対して，

$$R(\boldsymbol{X}, \boldsymbol{Y})\boldsymbol{Z} = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X \boldsymbol{Z} - \nabla_{[X,Y]} \boldsymbol{Z}$$

と定義し，曲率テンソル場とよぶ．これを第 1 の形式と呼ぶ．

また，

$$R(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{W}) : \mathfrak{X}(M) \times \mathfrak{X}(M) \times \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathbb{R}$$

を

$$R(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{W}) = \langle R(\boldsymbol{X}, \boldsymbol{Y})\boldsymbol{Z}, \boldsymbol{W} \rangle$$

と定義し，これも曲率テンソル場と呼ぶ．これを第 2 の形式と呼ぶ．第 1，第 2 形式とも $(1,3)$ 型（1 階反変，3 階共変）テンソルである[1]．

このノートでは主に第 1 形式の曲率テンソル場を用いる．

3.1.2　曲率テンソル場（第 1 の形式）の性質

上で定義した作用素（曲率テンソル場：第 1 の形式）について，以下が成り立つ．ただし，$\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \in \mathfrak{X}(M)$，$f \in C^\infty(M)$（$f$ はいくらでも微分可能）である．2)，3) は R がテンソル（多重線形写像）であることの帰結である．

1) $R(\boldsymbol{X}, \boldsymbol{Y})\boldsymbol{Z} = -R(\boldsymbol{Y}, \boldsymbol{X})\boldsymbol{Z}$

[1]第 1 形式はベクトル（場）を出力し，第 2 形式は実数を出力するが，第 1 形式の出力結果であるベクトルは，第 2 形式の出力結果を成分とするベクトルである．つまり，第 1 形式と第 2 形式の違いは，出力結果がベクトルの基底表示か成分表示かという点のみである．ちなみに，3 次元ユークリッド空間のベクトルの例では，$\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ を (x, y, z) 方向の基底ベクトルとして，ベクトル $\boldsymbol{x}$ を基底表示では $\boldsymbol{x} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$ と表し，成分表示では $\boldsymbol{x} = (x, y, z)$ と表す．

2) $R(f\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Z}=fR(\boldsymbol{Y},\boldsymbol{X})\boldsymbol{Z}$,　$R(\boldsymbol{X},f\boldsymbol{Y})\boldsymbol{Z}=fR(\boldsymbol{Y},\boldsymbol{X})\boldsymbol{Z}$

3) $R(\boldsymbol{X},\boldsymbol{Y})(f\boldsymbol{Z})=fR(\boldsymbol{Y},\boldsymbol{X})\boldsymbol{Z}$

(1) の証明：

定義から

$$R(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Z}=\nabla_X\nabla_Y\boldsymbol{Z}-\nabla_Y\nabla_X\boldsymbol{Z}-\nabla_{[\boldsymbol{X},\boldsymbol{Y}]}\boldsymbol{Z}$$

である．したがって，

$$R(\boldsymbol{Y},\boldsymbol{X})\boldsymbol{Z}=\nabla_Y\nabla_X\boldsymbol{Z}-\nabla_X\nabla_Y\boldsymbol{Z}-\nabla_{[\boldsymbol{Y},\boldsymbol{X}]}\boldsymbol{Z}$$

カッコ積は定義から $[\boldsymbol{X},\boldsymbol{Y}]=-[\boldsymbol{Y},\boldsymbol{X}]$ であるので，$R(\boldsymbol{Y},\boldsymbol{X})\boldsymbol{Z}=-R(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Z}$ が成り立つ．

(2) の証明：

$$R(f\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Z}=\nabla_{fX}\nabla_Y\boldsymbol{Z}-\nabla_Y\nabla_{fX}\boldsymbol{Z}-\nabla_{[fX,Y]}\boldsymbol{Z}$$

である．ここでまず，

$$[f\boldsymbol{X},\boldsymbol{Y}]=\nabla_{fX}\boldsymbol{Y}-\nabla_Y(f\boldsymbol{X})=f\nabla_X\boldsymbol{Y}-(\boldsymbol{Y}f)\boldsymbol{X}-f\nabla_Y\boldsymbol{X}=f[\boldsymbol{X},\boldsymbol{Y}]-(\boldsymbol{Y}f)\boldsymbol{X}$$

であり，したがって，

$$\nabla_{[fX,Y]}\boldsymbol{Z}=f\nabla_{[\boldsymbol{X},\boldsymbol{Y}]}\boldsymbol{Z}-(\boldsymbol{Y}f)\nabla_X\boldsymbol{Z}$$

となる．さらに，

$$\nabla_{fX}\nabla_Y\boldsymbol{Z}=f\nabla_X\nabla_Y\boldsymbol{Z}$$
$$\nabla_Y\nabla_{fX}\boldsymbol{Z}=\nabla_Y(f\nabla_X\boldsymbol{Z})=(\boldsymbol{Y}f)\nabla_X\boldsymbol{Z}+f\nabla_Y\nabla_X\boldsymbol{Z}$$

であるので，

$$\begin{aligned}R(f\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Z}&=f\nabla_X\nabla_Y\boldsymbol{Z}-\Big((\boldsymbol{Y}f)\nabla_X\boldsymbol{Z}+f\nabla_Y\nabla_X\boldsymbol{Z}\Big)-\Big(f\nabla_{[\boldsymbol{X},\boldsymbol{Y}]}\boldsymbol{Z}-(\boldsymbol{Y}f)\nabla_X\boldsymbol{Z}\Big)\\&=f\Big(\nabla_X\nabla_Y\boldsymbol{Z}-\nabla_Y\nabla_X\boldsymbol{Z}-\nabla_{[X,Y]}\boldsymbol{Z}\Big)=fR(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Z}\end{aligned}$$

(3) の証明：

$$R(\boldsymbol{X},\boldsymbol{Y})(f\boldsymbol{Z})=\nabla_X\nabla_Y(f\boldsymbol{Z})-\nabla_Y\nabla_X(f\boldsymbol{Z})-\nabla_{[X,Y]}(f\boldsymbol{Z})$$

であり，

$$\nabla_X\nabla_Y(f\boldsymbol{Z})=\nabla_X((\boldsymbol{Y}f)\boldsymbol{Z}+f\nabla_Y\boldsymbol{Z})=\boldsymbol{X}(\boldsymbol{Y}f)\boldsymbol{Z}+(\boldsymbol{Y}f)\nabla_X\boldsymbol{Z}+(\boldsymbol{X}f)\nabla_Y\boldsymbol{Z}+f\nabla_X\nabla_Y\boldsymbol{Z}$$

また，

$$\nabla_Y\nabla_X(f\boldsymbol{Z})=\nabla_Y((\boldsymbol{X}f)\boldsymbol{Z}+f\nabla_X\boldsymbol{Z})=\boldsymbol{Y}(\boldsymbol{X}f)\boldsymbol{Z}+(\boldsymbol{X}f)\nabla_Y\boldsymbol{Z}+(\boldsymbol{Y}f)\nabla_X\boldsymbol{Z}+f\nabla_Y\nabla_X\boldsymbol{Z}$$

したがって，

$$\nabla_X\nabla_Y(f\boldsymbol{Z})-\nabla_Y\nabla_X(f\boldsymbol{Z})=f\Big(\nabla_X\nabla_Y\boldsymbol{Z}-\nabla_Y\nabla_X\boldsymbol{Z}\Big)+\boldsymbol{X}(\boldsymbol{Y}f)\boldsymbol{Z}-\boldsymbol{Y}(\boldsymbol{X}f)\boldsymbol{Z}$$

また，

$$\nabla_{[X,Y]}(f\boldsymbol{Z})=([\boldsymbol{X},\boldsymbol{Y}]f)\boldsymbol{Z}+\nabla_{[X,Y]}\boldsymbol{Z}$$

であるので，したがって，

$$R(\boldsymbol{X},\boldsymbol{Y})(f\boldsymbol{Z})=fR(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Z}+\boldsymbol{X}(\boldsymbol{Y}f)\boldsymbol{Z}-\boldsymbol{Y}(\boldsymbol{X}f)\boldsymbol{Z}-([\boldsymbol{X},\boldsymbol{Y}]f)\boldsymbol{Z}$$

であるが，

$$([\boldsymbol{X},\boldsymbol{Y}]f)\boldsymbol{Z}=\Big[\boldsymbol{X}(\boldsymbol{Y}f)-\boldsymbol{Y}(\boldsymbol{X}f)\Big]\boldsymbol{Z}$$

であるので，結局，(3) が示せた．

3.1.3 ビアンキの恒等式

任意のベクトル場 $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \in \mathfrak{X}(M)$ に対して,

$$R(\boldsymbol{X}, \boldsymbol{Y})\boldsymbol{Z} + R(\boldsymbol{Y}, \boldsymbol{Z})\boldsymbol{X} + R(\boldsymbol{Z}, \boldsymbol{X})\boldsymbol{Y} = 0$$

が成り立つ.

証明：

$$\begin{aligned} R(\boldsymbol{X}, \boldsymbol{Y})\boldsymbol{Z} &= \nabla_X \nabla_Y \boldsymbol{Z} - \nabla_Y \nabla_X \boldsymbol{Z} - \nabla_{[X,Y]}\boldsymbol{Z} \\ R(\boldsymbol{Y}, \boldsymbol{Z})\boldsymbol{X} &= \nabla_Y \nabla_Z \boldsymbol{X} - \nabla_Z \nabla_Y \boldsymbol{X} - \nabla_{[Y,Z]}\boldsymbol{X} \\ R(\boldsymbol{Z}, \boldsymbol{X})\boldsymbol{Y} &= \nabla_Z \nabla_X \boldsymbol{Y} - \nabla_X \nabla_Z \boldsymbol{Y} - \nabla_{[Z,X]}\boldsymbol{Y} \end{aligned}$$

であるので，全部足せば右辺は

$$\begin{aligned} 右辺 = \nabla_X \nabla_Y \boldsymbol{Z} - \nabla_Y \nabla_X \boldsymbol{Z} + \nabla_Y \nabla_Z \boldsymbol{X} - \nabla_Z \nabla_Y \boldsymbol{X} + \nabla_Z \nabla_X \boldsymbol{Y} - \nabla_X \nabla_Z \boldsymbol{Y} \\ - \nabla_{[X,Y]}\boldsymbol{Z} - \nabla_{[Y,Z]}\boldsymbol{X} - \nabla_{[Z,X]}\boldsymbol{Y} \\ = \nabla_X(\nabla_Y \boldsymbol{Z} - \nabla_Z \boldsymbol{Y}) + \nabla_Y(\nabla_Z \boldsymbol{X} - \nabla_X \boldsymbol{Z}) + \nabla_Z(\nabla_X \boldsymbol{Y} - \nabla_Y \boldsymbol{X}) \\ - \nabla_{[X,Y]}\boldsymbol{Z} - \nabla_{[Y,Z]}\boldsymbol{X} - \nabla_{[Z,X]}\boldsymbol{Y} \end{aligned}$$

ここで,

$$\begin{aligned} \nabla_Y \boldsymbol{Z} - \nabla_Z \boldsymbol{Y} &= [\boldsymbol{Y}, \boldsymbol{Z}] \\ \nabla_Z \boldsymbol{X} - \nabla_X \boldsymbol{Z} &= [\boldsymbol{Z}, \boldsymbol{X}] \\ \nabla_X \boldsymbol{Y} - \nabla_Y \boldsymbol{X} &= [\boldsymbol{X}, \boldsymbol{Y}] \end{aligned}$$

であるので，結局,

$$右辺 = \nabla_X[\boldsymbol{Y}, \boldsymbol{Z}] - \nabla_{[Y,Z]}\boldsymbol{X} + \nabla_Y[\boldsymbol{Z}, \boldsymbol{X}] - -\nabla_{[Z,X]}\boldsymbol{Y} + \nabla_Z[\boldsymbol{X}, \boldsymbol{Y}] - \nabla_{[X,Y]}\boldsymbol{Z}$$

であるが，さらに,

$$\begin{aligned} \nabla_X[\boldsymbol{Y}, \boldsymbol{Z}] - \nabla_{[Y,Z]}\boldsymbol{X} &= [\boldsymbol{X}, [\boldsymbol{Y}, \boldsymbol{Z}]] \\ \nabla_Y[\boldsymbol{Z}, \boldsymbol{X}] - \nabla_{[Z,X]}\boldsymbol{Y} &= [\boldsymbol{Y}, [\boldsymbol{Z}, \boldsymbol{X}]] \\ \nabla_Z[\boldsymbol{X}, \boldsymbol{Y}] - \nabla_{[X,Y]}\boldsymbol{Z} &= [\boldsymbol{Z}, [\boldsymbol{X}, \boldsymbol{Y}]] \end{aligned}$$

であるので，カッコ積のヤコビ恒等式から

$$[\boldsymbol{X}, [\boldsymbol{Y}, \boldsymbol{Z}]] + [\boldsymbol{Y}, [\boldsymbol{Z}, \boldsymbol{X}]] + [\boldsymbol{Z}, [\boldsymbol{X}, \boldsymbol{Y}]] = 0$$

である．したがって,

$$R(\boldsymbol{X}, \boldsymbol{Y})\boldsymbol{Z} + R(\boldsymbol{Y}, \boldsymbol{Z})\boldsymbol{X} + R(\boldsymbol{Z}, \boldsymbol{X})\boldsymbol{Y} = 0$$

を得る.

3.1.4 曲率テンソルの成分表示

局所座標系を用いて曲率テンソル $R(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Z}$ を成分で表す．曲率テンソルは，基底表示をすると

$$R(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Z} = R^{\ell}_{ijk} dx^k \otimes dx^j \otimes dx^i \otimes \frac{\partial}{\partial x^\ell}$$

であり，テンソルの局所座標系 (x^i) における成分 R^{ℓ}_{ijk} は

$$R(\frac{\partial}{\partial u^k}, \frac{\partial}{\partial u^j})\frac{\partial}{\partial u^i} = R^{\ell}_{ijk}\frac{\partial}{\partial u^\ell}$$

として求まる．上式左辺を計算する．

まず,

$$R(\frac{\partial}{\partial u^k}, \frac{\partial}{\partial u^j})\frac{\partial}{\partial u^i} = \nabla_{\frac{\partial}{\partial u^k}}\nabla_{\frac{\partial}{\partial u^j}}\frac{\partial}{\partial u^i} - \nabla_{\frac{\partial}{\partial u^j}}\nabla_{\frac{\partial}{\partial u^k}}\frac{\partial}{\partial u^i} - \nabla_{[\frac{\partial}{\partial u^k}, \frac{\partial}{\partial u^j}]}\frac{\partial}{\partial u^i}$$

を計算する[2]．ここで，$[\frac{\partial}{\partial u^k}, \frac{\partial}{\partial u^j}] = 0$ であるので,

$$R(\frac{\partial}{\partial u^k}, \frac{\partial}{\partial u^j})\frac{\partial}{\partial u^i} = \nabla_{\frac{\partial}{\partial u^k}}\nabla_{\frac{\partial}{\partial u^j}}\frac{\partial}{\partial u^i} - \nabla_{\frac{\partial}{\partial u^j}}\nabla_{\frac{\partial}{\partial u^k}}\frac{\partial}{\partial u^i}$$

となって，曲率テンソル $R(\frac{\partial}{\partial u^k}, \frac{\partial}{\partial u^j})\frac{\partial}{\partial u^i}$ は共変微分の順序を変えた差に等しくなる．

ここで,

$$\nabla_{\frac{\partial}{\partial u^i}}\frac{\partial}{\partial u^j} = \sum_{\ell}\Gamma^{\ell}{}_{ij}\frac{\partial}{\partial u^\ell} = \Gamma^{\ell}{}_{ij}\frac{\partial}{\partial u^\ell}$$

を用いる．この式は，接空間の j 方向への基底 $\frac{\partial}{\partial u^j}$ の i 方向への共変微分を，$\Gamma^{k}{}_{ij}$ を係数として基底 $\frac{\partial}{\partial u^k}$ で展開したものである．すると,

$$\begin{aligned}
\nabla_{\frac{\partial}{\partial u^k}}(\nabla_{\frac{\partial}{\partial u^j}}\frac{\partial}{\partial u^i}) = \nabla_{\frac{\partial}{\partial u^k}}(\Gamma^{\ell}{}_{ji}\frac{\partial}{\partial u^\ell}) &= \frac{\partial\Gamma^{\ell}{}_{ji}}{\partial u^k}\frac{\partial}{\partial u^\ell} + \Gamma^{\ell}{}_{ji}\nabla_{\frac{\partial}{\partial u^k}}\frac{\partial}{\partial u^\ell} \\
&= \frac{\partial\Gamma^{\ell}{}_{ji}}{\partial u^k}\frac{\partial}{\partial u^\ell} + \Gamma^{\ell}{}_{ji}\Gamma^{m}{}_{k\ell}\frac{\partial}{\partial u^m} = \frac{\partial\Gamma^{\ell}{}_{ij}}{\partial u^k}\frac{\partial}{\partial u^\ell} + \Gamma^{m}{}_{ij}\Gamma^{\ell}{}_{km}\frac{\partial}{\partial u^\ell}
\end{aligned}$$

また,

$$\nabla_{\frac{\partial}{\partial u^j}}\nabla_{\frac{\partial}{\partial u^k}}\frac{\partial}{\partial u^i} = \frac{\partial\Gamma^{\ell}{}_{ik}}{\partial u^j}\frac{\partial}{\partial u^\ell} + \Gamma^{m}{}_{ik}\Gamma^{\ell}{}_{jm}\frac{\partial}{\partial u^\ell}$$

したがって,

$$\begin{aligned}
R(\frac{\partial}{\partial u^k}, \frac{\partial}{\partial u^j})\frac{\partial}{\partial u^i} &= \frac{\partial\Gamma^{\ell}{}_{ji}}{\partial u^k}\frac{\partial}{\partial u^\ell} + \Gamma^{m}{}_{ji}\Gamma^{\ell}{}_{im}\frac{\partial}{\partial u^\ell} - \left(\frac{\partial\Gamma^{\ell}{}_{ik}}{\partial u^j}\frac{\partial}{\partial u^\ell} + \Gamma^{m}{}_{ik}\Gamma^{\ell}{}_{jm}\frac{\partial}{\partial u^\ell}\right) \\
&= \left(\frac{\partial\Gamma^{\ell}{}_{ji}}{\partial u^k} - \frac{\partial\Gamma^{\ell}{}_{ik}}{\partial u^j} + \Gamma^{m}{}_{ji}\Gamma^{\ell}{}_{km} - \Gamma^{m}{}_{ik}\Gamma^{\ell}{}_{jm}\right)\frac{\partial}{\partial u^\ell} \\
&= \left(\frac{\partial\Gamma^{n}{}_{ij}}{\partial u^k} - \frac{\partial\Gamma^{n}{}_{ik}}{\partial u^j} + \Gamma^{m}{}_{ij}\Gamma^{n}{}_{km} - \Gamma^{m}{}_{ik}\Gamma^{n}{}_{jm}\right)\frac{\partial}{\partial u^n} \qquad (3.1)
\end{aligned}$$

---

[2] R^{ℓ}_{ijk} に関して，教科書にある公式を導くには微分の順序に注意しなければならない．すなわち微分の順序を $\boldsymbol{p}_{ijk}$ と同じにしなければならない．つまり,

$$\boldsymbol{p}_{ijk} = \frac{\partial}{\partial u^k}\left(\frac{\partial}{\partial u^j}\left(\frac{\partial}{\partial u^i}\boldsymbol{p}\right)\right)$$

であるので，$i \to j \to k$ の順番で微分をするように添え字を決める．

を得る．これを，

$$R(\frac{\partial}{\partial u^k}, \frac{\partial}{\partial u^j})\frac{\partial}{\partial u^i} = R^{\ell}_{ijk}\frac{\partial}{\partial u^\ell}$$

とする．したがって，

$$R^{\ell}_{ijk} = \frac{\partial \Gamma^{\ell}{}_{ij}}{\partial u^k} - \frac{\partial \Gamma^{\ell}{}_{ik}}{\partial u^j} + \Gamma^{m}{}_{ij}\,\Gamma^{\ell}{}_{km} - \Gamma^{m}{}_{ik}\,\Gamma^{\ell}{}_{jm} \tag{3.2}$$

を得る[3]．

3.1.5 捩率テンソル場

$$T\,:\,\mathfrak{X}(M)\times\mathfrak{X}(M)\to\mathfrak{X}(M)$$

に対して，

$$T(\boldsymbol{X},\boldsymbol{Y}) = \nabla_X\boldsymbol{Y} - \nabla_Y\boldsymbol{X} - [\boldsymbol{X},\boldsymbol{Y}]$$

と定義し，捩率テンソル場とよぶ．T が双線形写像であること，つまり，

$$T(f\boldsymbol{X}, g\boldsymbol{Y}) = fgT(\boldsymbol{X},\boldsymbol{Y})$$

が成り立つことは簡単に示せる．局所座標 $(x^1, x^2, \ldots, x^n)$ で表すと，

$$T(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}) = \nabla_{\frac{\partial}{\partial x^i}}\frac{\partial}{\partial x^j} - \nabla_{\frac{\partial}{\partial x^j}}\frac{\partial}{\partial x^i} - [\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}] = \Gamma^{\ell}{}_{ij}\frac{\partial}{\partial x^\ell} - \Gamma^{\ell}{}_{ji}\frac{\partial}{\partial x^\ell} = \left(\Gamma^{\ell}{}_{ij} - \Gamma^{\ell}{}_{ji}\right)\frac{\partial}{\partial x^\ell}$$

したがって，$T(\boldsymbol{X},\boldsymbol{Y})$ は 1 階反変 2 階共変テンソルである．

$$T(\boldsymbol{X},\boldsymbol{Y}) = T^k_{ij}\frac{\partial}{\partial x^k}$$

とすれば，

$$T^k_{ij} = \Gamma^{k}{}_{ij} - \Gamma^{k}{}_{ji} \tag{3.3}$$

である．つまり，$T=0$ の条件は接続係数 $\Gamma^{k}{}_{ij}$ が下付きの添え字に関して対称となることと同値である．

3.2 平坦な多様体とアフィン座標系

3.2.1 平坦な多様体

平坦性の定義　アフィン接続を持つ多様体 M において，曲率テンソル場 R も捩率テンソル場 T もともに恒等的にゼロとなるとき，M は平坦であると言う．

アフィン座標系：定義　アフィン接続 ∇ を持つ多様体 M の局所座標系 $(x^1, x^2, \ldots, x^n)$ において，接続係数 $\Gamma^{k}{}_{ij}$ が全て恒等的にゼロとなるとき，この座標系を $\nabla-$ アフィン座標系と呼ぶ．

定理　アフィン接続 ∇ を持つ多様体 M において，次の 2 条件は同値である．

[3]多くの本では，共変微分の順序を変えた場合のちがいから上式を導いている．この場合，共変微分の式でクリストッフェル記号を含む項はマイナス記号が付くため正負が上式と逆になっている．

(i) M は平坦である．

(ii) M の各点の周りにアフィン座標系が存在する．

証明：　(ii)→(i)

この部分は簡単である．アフィン座標系においては接続係数 ${\Gamma^k}_{ij}$ は全てゼロであるので，このとき，曲率テンソル R も捩率テンソル T もゼロとなる．

証明：　(i)→(ii)

まず，接続係数の変換測を導く．

局所座標系 x^i において[4]，x^j 方向の接ベクトルの x^i 方向への方向微分は,

$$\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = {\Gamma^k}_{ij} \frac{\partial}{\partial x^k}$$

と書ける．また，別の局所座標系 ξ^a を用いると，

$$\nabla_{\frac{\partial}{\partial \xi^a}} \frac{\partial}{\partial \xi^b} = {\Gamma^c}_{ab} \frac{\partial}{\partial \xi^c}$$

と書ける．2 つの座標系で表すことのできる領域では，接続係数 ${\Gamma^k}_{ij}$ と ${\Gamma^c}_{ab}$ の間に何らかの変換測があるはずである．この変換測を導く．

そのため，$\nabla_{\frac{\partial}{\partial x^i}} \dfrac{\partial}{\partial x^j}$ を 2 通りで計算する．まず,

$$\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = {\Gamma^k}_{ij} \frac{\partial}{\partial x^k} = {\Gamma^k}_{ij} \frac{\partial \xi^c}{\partial x^k} \frac{\partial}{\partial \xi^c} \tag{3.4}$$

と表せる．上式は，接ベクトル変換測：$\dfrac{\partial}{\partial x^k} = \dfrac{\partial \xi^c}{\partial x^k} \dfrac{\partial}{\partial \xi^c}$ を用いただけである[5]．

一方，この変換測を用いて，

$$\begin{aligned}
\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} &= \nabla_{\frac{\partial}{\partial x^i}} \left(\frac{\partial \xi^b}{\partial x^j} \frac{\partial}{\partial \xi^b} \right) = \frac{\partial^2 \xi^b}{\partial x^i \partial x^j} \frac{\partial}{\partial \xi^b} + \frac{\partial \xi^b}{\partial x^j} \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial \xi^b} = \frac{\partial^2 \xi^b}{\partial x^i \partial x^j} \frac{\partial}{\partial \xi^b} + \frac{\partial \xi^b}{\partial x^j} \nabla_{\frac{\partial \xi^a}{\partial x^i} \frac{\partial}{\partial \xi^a}} \frac{\partial}{\partial \xi^b} \\
&= \frac{\partial^2 \xi^b}{\partial x^i \partial x^j} \frac{\partial}{\partial \xi^b} + \frac{\partial \xi^b}{\partial x^j} \frac{\partial \xi^a}{\partial x^i} \nabla_{\frac{\partial}{\partial \xi^a}} \frac{\partial}{\partial \xi^b} = \frac{\partial^2 \xi^b}{\partial x^i \partial x^j} \frac{\partial}{\partial \xi^b} + \frac{\partial \xi^b}{\partial x^j} \frac{\partial \xi^a}{\partial x^i} {\Gamma^c}_{ab} \frac{\partial}{\partial \xi^c} \\
&= \left(\frac{\partial^2 \xi^c}{\partial x^i \partial x^j} + \frac{\partial \xi^b}{\partial x^j} \frac{\partial \xi^a}{\partial x^i} {\Gamma^c}_{ab} \right) \frac{\partial}{\partial \xi^c}
\end{aligned} \tag{3.5}$$

である．ここで，式 (3.4) と (3.5) を比較すれば,

$${\Gamma^k}_{ij} \frac{\partial \xi^c}{\partial x^k} = \frac{\partial^2 \xi^c}{\partial x^i \partial x^j} + \frac{\partial \xi^b}{\partial x^j} \frac{\partial \xi^a}{\partial x^i} {\Gamma^c}_{ab}$$

[4]局所座標系 $x^1, x^2, \ldots, x^n$ と書くべきところを省略してこのように書いた．

[5]この変換測は，より明示的に書くと，

$$\begin{bmatrix} \dfrac{\partial}{\partial x^1} \\ \vdots \\ \dfrac{\partial}{\partial x^n} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial \xi^1}{\partial x^1} & \cdots & \dfrac{\partial \xi^n}{\partial x^1} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial \xi^1}{\partial x^n} & \cdots & \dfrac{\partial \xi^n}{\partial x^n} \end{bmatrix} \begin{bmatrix} \dfrac{\partial}{\partial \xi^1} \\ \vdots \\ \dfrac{\partial}{\partial \xi^n} \end{bmatrix}$$

となる．接ベクトルにヤコビ行列をかけて別の座標系の接ベクトルを得ている．

を得る．ヤコビ行列 $\dfrac{\partial \xi^c}{\partial x^k}$ の逆行列を両辺に乗じて

$$\Gamma^k{}_{ij} = \frac{\partial^2 \xi^c}{\partial x^i \partial x^j}\frac{\partial x^k}{\partial \xi^c} + \frac{\partial \xi^b}{\partial x^j}\frac{\partial \xi^a}{\partial x^i}\frac{\partial x^k}{\partial \xi^c}\Gamma^c{}_{ab}$$

を得る．上式が，座標変換 $(\xi^1, \ldots, \xi^n) \to (x^1, \ldots, x^n)$ に関する接続係数の座標変換測である．

ここから，(i)→(ii) の証明に入る．任意の局所座標系 $x^1, \ldots, x^n$ に対して，ある座標系 $\xi^1, \ldots, \xi^n$ が存在して，その座標系における接続係数はゼロとなるので，

$$0 = \Gamma^c{}_{ab} = \frac{\partial^2 x^\ell}{\partial \xi^a \partial \xi^b}\frac{\partial \xi^c}{\partial x^\ell} + \frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b}\frac{\partial \xi^c}{\partial x^k}\Gamma^k{}_{ij} \tag{3.6}$$

となる．以下では，このような座標系 (ξ^a) が，任意の座標系 (x^i) に対して必ず存在することを示す．このため，まず，恒等式

$$\frac{\partial x^\ell}{\partial \xi^b}\frac{\partial \xi^b}{\partial x^j} = \delta^\ell_j$$

の両辺を x^i で偏微分する．

$$\frac{\partial^2 x^\ell}{\partial \xi^a \partial \xi^b}\frac{\partial \xi^a}{\partial x^i}\frac{\partial \xi^b}{\partial x^j} + \frac{\partial x^\ell}{\partial \xi^b}\frac{\partial^2 \xi^b}{\partial x^i \partial x^j} = 0$$

であるので，両辺にヤコビ行列 $\dfrac{\partial \xi^a}{\partial x^i}$ と $\dfrac{\partial \xi^b}{\partial x^j}$ の逆行列をかけると，

$$\frac{\partial^2 x^\ell}{\partial \xi^a \partial \xi^b} = -\frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b}\frac{\partial x^\ell}{\partial \xi^d}\frac{\partial^2 \xi^d}{\partial x^i \partial x^j}$$

を得る．式 (3.6) に代入して，

$$\begin{aligned} 0 = \Gamma^c{}_{ab} &= \frac{\partial^2 x^\ell}{\partial \xi^a \partial \xi^b}\frac{\partial \xi^c}{\partial x^\ell} + \frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b}\frac{\partial \xi^c}{\partial x^k}\Gamma^k{}_{ij} = -\frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b}\frac{\partial x^\ell}{\partial \xi^d}\frac{\partial^2 \xi^d}{\partial x^i \partial x^j}\frac{\partial \xi^c}{\partial x^\ell} + \frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b}\frac{\partial \xi^c}{\partial x^k}\Gamma^k{}_{ij} \\ &= \frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b}\frac{\partial \xi^c}{\partial x^k}\Gamma^k{}_{ij} - \frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b}\frac{\partial^2 \xi^c}{\partial x^i \partial x^j} = \frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b}\left(\frac{\partial \xi^c}{\partial x^k}\Gamma^k{}_{ij} - \frac{\partial^2 \xi^c}{\partial x^i \partial x^j}\right) \end{aligned} \tag{3.7}$$

を得る．したがって，$\Gamma^c{}_{ab} = 0$ が成り立つのは，

$$\frac{\partial \xi^c}{\partial x^k}\Gamma^k{}_{ij} = \frac{\partial^2 \xi^c}{\partial x^i \partial x^j} \tag{3.8}$$

が成り立つのと同値である．上式は $\xi^1, \ldots, \xi^n$ に関する 2 階の偏微分方程式である．この方程式が解 $\xi^1, \ldots, \xi^n$ を持てばこの座標系における $\Gamma^c{}_{ab}$ は恒等的にゼロとなる．

以下では，条件 (i)，つまり曲率テンソル $R = 0$ と捩率テンソル $T = 0$ の条件下で，式 (3.8) が（任意の座標系 (x^i) に対して）必ず解を持つことを示す．これは，偏微分方程式の解の存在条件に関する議論である．

式 (3.8) を，2 つに分けて書くと

$$\frac{\partial \xi^c}{\partial x^k} = \theta^c_k$$
$$\frac{\partial \theta^c_j}{\partial x^i} = \theta^c_k \Gamma^k{}_{ij}$$

との，連立微分方程式となる．この連立微分方程式が解を持つ条件（可積分条件）は

$$\frac{\partial^2 \xi^c}{\partial x^i \partial x^j} = \frac{\partial^2 \xi^c}{\partial x^j \partial x^i} \tag{3.9}$$

$$\frac{\partial^2 \theta^c_k}{\partial x^i \partial x^j} = \frac{\partial^2 \theta^c_k}{\partial x^j \partial x^i} \tag{3.10}$$

である．まず，式 (3.9) は

$$\frac{\partial^2 \xi^c}{\partial x^i \partial x^j} = \frac{\partial^2 \xi^c}{\partial x^j \partial x^i} \Longleftrightarrow \frac{\partial}{\partial x^i}\theta^c_j = \frac{\partial}{\partial x^j}\theta^c_i \Longleftrightarrow \theta^c_k {\Gamma^k}_{ij} = \theta^c_k {\Gamma^k}_{ji} \Longleftrightarrow \theta^c_k T^k_{ij} = 0$$

となる．ここで，捩率テンソル $T^k_{ij} = 0$ であるので，この可積分条件は成り立つ．次に，式 (3.10) は,

$$\begin{aligned}
\frac{\partial^2 \theta^c_k}{\partial x^i \partial x^j} = \frac{\partial^2 \theta^c_k}{\partial x^j \partial x^i} &\Longleftrightarrow \frac{\partial}{\partial x^i}(\theta^c_\ell {\Gamma^\ell}_{jk}) = \frac{\partial}{\partial x^j}(\theta^c_\ell {\Gamma^\ell}_{ik}) \\
&\Longleftrightarrow \frac{\partial \theta^c_\ell}{\partial x^i}{\Gamma^\ell}_{jk} + \theta^c_\ell \frac{\partial {\Gamma^\ell}_{jk}}{\partial x^i} - \frac{\partial \theta^c_\ell}{\partial x^j}{\Gamma^\ell}_{ik} - \theta^c_\ell \frac{\partial {\Gamma^\ell}_{ik}}{\partial x^j} = 0 \\
&\Longleftrightarrow \theta^c_m {\Gamma^m}_{i\ell}{\Gamma^\ell}_{jk} + \theta^c_m \frac{\partial {\Gamma^m}_{jk}}{\partial x^i} - \theta^c_m {\Gamma^m}_{j\ell}{\Gamma^\ell}_{ik} - \theta^c_m \frac{\partial {\Gamma^m}_{ik}}{\partial x^j} = 0 \\
&\Longleftrightarrow \theta^c_m \left({\Gamma^m}_{i\ell}{\Gamma^\ell}_{jk} + \frac{\partial {\Gamma^m}_{jk}}{\partial x^i} - {\Gamma^m}_{j\ell}{\Gamma^\ell}_{ik} - \frac{\partial {\Gamma^m}_{ik}}{\partial x^j} \right) = 0 \Longleftrightarrow \theta^c_m R^m_{ijk} = 0
\end{aligned}$$

となる．リーマン曲率テンソル $R^m_{ijk} = 0$ であるので，この可積分条件も成り立つ．したがって，式 (3.8) に示す微分方程式は解 $\xi^1, \ldots, \xi^n$ を持ち，この解のもとで接続係数は全てゼロとなる．

定理 M に局所座標系 $(x^1, \ldots, x^n)$ と $(\xi^1, \ldots, \xi^n)$ があって，(x^i) はアフィン座標系であったとする．このとき，(ξ^a) もアフィン座標系であるための必要十分条件は座標変換がアフィン変換（線形変換）で書けること，すなわち，

$$\begin{bmatrix} x^1 \\ \vdots \\ x^n \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} \xi^1 \\ \vdots \\ \xi^n \end{bmatrix} + \boldsymbol{b}$$

と書けることである．

証明： 証明は簡単である．接続係数の座標変換は

$${\Gamma^k}_{ij} = \frac{\partial^2 \xi^c}{\partial x^i \partial x^j}\frac{\partial x^k}{\partial \xi^c} + \frac{\partial \xi^b}{\partial x^j}\frac{\partial \xi^a}{\partial x^i}\frac{\partial x^k}{\partial \xi^c}{\Gamma^c}_{ab}$$

と表される．${\Gamma^k}_{ij} = 0$ のもとで ${\Gamma^c}_{ab} = 0$ となるための必要十分条件は

$$\frac{\partial^2 \xi^c}{\partial x^i \partial x^j} = 0$$

となることであり，これは，座標変換がアフィン変換（座標変換が線形変換 $\xi^c = ax^i + b$）であることと同値である．

3.2.2 Riemann 接続

定義： M 上で定義された $(0, 2)$ 型テンソル場 g であって，各点 $p \in M$ において g_p が $T_p(M)$ 上の内積であるものを M のリーマン計量と言う．局所座標系を用いて，

$$g = g_{ij} dx^i \otimes dx^j$$

と表すと

$$g\left(\frac{\partial}{\partial x^m}, \frac{\partial}{\partial x^n}\right) = g_{ij}dx^i(\frac{\partial}{\partial x^m})dx^j(\frac{\partial}{\partial x^n}) = g_{ij}\delta^i_m\delta^j_n = g_{mn}$$

を得る．

定義：リーマン計量 g が与えられた多様体のことをリーマン多様体という．以下，リーマン多様体を (M, g) で表す．

以下，接ベクトル $\dfrac{\partial}{\partial x^i}$ を

$$\partial_i = \frac{\partial}{\partial x^i}$$

との表記法であらわす．

定義： リーマン多様体の座標近傍 $(x^1, \ldots, x^n)$ において，

$$\Gamma_{ij,k} = \Gamma^\ell{}_{ij}\, g_{\ell k} = \frac{1}{2}\left(\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij}\right) \tag{3.11}$$

で定まる M のアフィン接続を Riemann 接続，あるいは，Levi-Civita 接続という．ここで，$\Gamma_{ij,k}$ は，接ベクトル $\nabla_{\partial_i}\partial_j$ と ∂_k の内積である．実際，

$$\Gamma_{ij,k} = g\left(\nabla_{\partial_i}\partial_j, \partial_k\right) = g\left(\Gamma^\ell{}_{ij}\,\partial_\ell, \partial_k\right) = \Gamma^\ell{}_{ij}\, g\left(\partial_\ell, \partial_k\right) = \Gamma^\ell{}_{ij}\, g_{\ell k}$$

g_{ij} は正定値行列なので逆行列が存在する．それを g^{ij} と書くことにすれば

$$g_{ij}g^{jk} = \delta^k_i$$

が成り立つ．したがって，

$$\Gamma_{ij,k}g^{km} = \Gamma^\ell{}_{ij}\, g_{\ell k}g^{km} = \Gamma^\ell{}_{ij}\,\delta^m_\ell = \Gamma^m{}_{ij}$$

である．

リーマン接続は次の著しい性質を持つ．

定理　なめらかな曲線 $C = \{p(t); a \geq t \geq b\}$ に沿った LC(Levi-Civita) 平行移動を Π^a_b と書くと，任意の $\boldsymbol{v}, \boldsymbol{w} \in T_{p(a)}M$ に対して，

$$g_{p(b)}\left(\Pi^a_b\boldsymbol{v}, \Pi^a_b\boldsymbol{w}\right) = g_{p(a)}(\boldsymbol{v}, \boldsymbol{w})$$

が成り立つ．すなわち，内積は Levi-Civita 平行移動に対して不変である．

証明：　C に沿って平行な 2 つのベクトル場 $\boldsymbol{Y} = \{\boldsymbol{Y}_{p(t)}\}$ と $\boldsymbol{Z} = \{\boldsymbol{Z}_{p(t)}\}$ に対して，内積 $g_{p(t)}(\boldsymbol{Y}_{p(t)}, \boldsymbol{Z}_{p(t)})$ が不変であること，すなわち，

$$\frac{d}{dt}g_{p(t)}(\boldsymbol{Y}_{p(t)}, \boldsymbol{Z}_{p(t)}) = 0$$

を示す．

局所座標系 (x^i) を用いて成分表示すれば，

$$\boldsymbol{Y}_{p(t)} = Y^i(t)\left(\frac{\partial}{\partial x^i}\right)_{p(t)} \qquad \boldsymbol{Z}_{p(t)} = Z^i(t)\left(\frac{\partial}{\partial x^i}\right)_{p(t)}$$

である．すると，

$$\begin{aligned}\frac{d}{dt}g_{p(t)}(\boldsymbol{Y}_{p(t)}, \boldsymbol{Z}_{p(t)}) = \frac{d}{dt}\left[g_{ij}Y^i(t)Z^j(t)\right] &= \frac{dg_{ij}}{dt}Y^i(t)Z^j(t) + g_{ij}\dot{Y}^i(t)Z^j(t) + g_{ij}Y^i(t)\dot{Z}^j(t) \\ &= \frac{\partial g_{ij}}{\partial x^k}\dot{x}^k Y^i(t)Z^j(t) + g_{ij}\dot{Y}^i(t)Z^j(t) + g_{ij}Y^i(t)\dot{Z}^j(t) \end{aligned} \tag{3.12}$$

が成り立つ．ここで，$p(t)$ に沿って平行なベクトル場 $\boldsymbol{X}$ について

$$\frac{dX^k}{dt} + X^j \frac{dx^i}{dt} \Gamma^k{}_{ij} = 0 \quad \rightarrow \quad \dot{X}^k = -X^j \dot{x}^i \Gamma^k{}_{ij}$$

が成り立つ．したがって，

$$\dot{Y}^i = -\Gamma^i{}_{k\ell}\, \dot{x}^k Y^\ell \qquad \dot{Z}^j = -\Gamma^j{}_{k\ell}\, \dot{x}^k Z^\ell$$

が成り立つので，これらを式 (3.12) に代入して，

$$\begin{aligned}
\frac{d}{dt} g_{p(t)}(\boldsymbol{Y}_{p(t)}, \boldsymbol{Z}_{p(t)}) &= \frac{d}{dt}\left[g_{ij} Y^i(t) Z^j(t)\right] = \frac{\partial g_{ij}}{\partial x^k} \dot{x}^k Y^i(t) Z^j(t) + g_{ij} \dot{Y}^i(t) Z^j(t) + g_{ij} Y^i(t) \dot{Z}^j(t) \\
&= \frac{\partial g_{ij}}{\partial x^k} \dot{x}^k Y^i Z^j - g_{ij}(\Gamma^i{}_{k\ell}\, \dot{x}^k Y^\ell) Z^j - g_{ij} Y^i (\Gamma^j{}_{k\ell}\, \dot{x}^k Z^\ell) \\
&= \frac{\partial g_{ij}}{\partial x^k} \dot{x}^k Y^i Z^j - g_{mj} \Gamma^m{}_{ki}\, \dot{x}^k Y^i Z^j - g_{im} \Gamma^m{}_{kj}\, \dot{x}^k Y^i Z^j = \left(\frac{\partial g_{ij}}{\partial x^k} - g_{mj} \Gamma^m{}_{ki} - g_{im} \Gamma^m{}_{kj} \right) \dot{x}^k Y^i Z^j \\
&= \left(\frac{\partial g_{ij}}{\partial x^k} - \Gamma_{ki,j} - \Gamma_{kj,i} \right) \dot{x}^k Y^i Z^j
\end{aligned} \tag{3.13}$$

を得る．

ところで，リーマン接続の定義式 (3.11)：

$$\Gamma_{ij,k} = \frac{1}{2}\left(\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij}\right)$$

から，

$$\Gamma_{ki,j} = \frac{1}{2}\left(\partial_k g_{ij} + \partial_i g_{jk} - \partial_j g_{ki}\right) \qquad \Gamma_{kj,i} = \Gamma_{jk,i} = \frac{1}{2}\left(\partial_j g_{ki} + \partial_k g_{ij} - \partial_i g_{jk}\right)$$

を得る．したがって，

$$\frac{\partial g_{ij}}{\partial x^k} - \Gamma_{ki,j} - \Gamma_{kj,i} = \partial_k g_{ij} - \frac{1}{2}\left(\partial_k g_{ij} + \partial_i g_{jk} - \partial_j g_{ki}\right) - \frac{1}{2}\left(\partial_j g_{ki} + \partial_k g_{ij} - \partial_i g_{jk}\right) = 0$$

となる．したがって，リーマン接続が与えられたとき，任意の曲線に沿って平行なベクトル場どうしの内積は不変に保たれる．**証明終**

アフィン接続 ∇ を持つ Riemann 多様体 (M, g) において，その計量が平行移動で保たれる場合に，接続 ∇ は計量的であるという．上の証明からわかるように接続が計量的である必要十分条件は

$$\partial_k g_{ij} - \Gamma_{ki,j} - \Gamma_{kj,i} = 0$$

が成り立つことであり，座標系に依らない形で書けば

$$\boldsymbol{X} g(\boldsymbol{Y}, \boldsymbol{Z}) = g(\nabla_X \boldsymbol{Y}, \boldsymbol{Z}) + g(\boldsymbol{Y}, \nabla_X \boldsymbol{Z})$$

となる．**上の関係式は，内積に対するライプニッツ則に他ならない．**

証明：

$$\boldsymbol{X} = \partial_k \qquad \boldsymbol{Y} = \partial_i \qquad \boldsymbol{Z} = \partial_j$$

とすれば,

$$\nabla_X \boldsymbol{Y} = \nabla_{\partial_k}\partial_i = \Gamma^{\ell}{}_{ki}\,\partial_\ell$$

であり,

$$g(\nabla_X \boldsymbol{Y}, \boldsymbol{Z}) = \Gamma^{\ell}{}_{ki}\,g(\partial_\ell, \partial_j) = \Gamma_{ki,j}$$

となる，また,

$$\nabla_X \boldsymbol{Z} = \nabla_{\partial_k}\partial_j = \Gamma^{\ell}{}_{kj}\,\partial_\ell$$

したがって，であるので,

$$g(\boldsymbol{Y}, \nabla_X \boldsymbol{Z}) = \Gamma^{\ell}{}_{kj}\,g(\partial_i, \partial_e ll) = \Gamma_{kj,i}$$

また,

$$\boldsymbol{X} g(\boldsymbol{Y}, \boldsymbol{Z}) = \partial_k g_{ij}$$

であるので,

$$\partial_k g_{ij} = \Gamma_{ki,j} + \Gamma_{kj,i}$$

を得る．

定理 リーマン多様体 (M, g) のアフィン接続がリーマン接続であるための必要十分条件は接続が計量的であって捩率 T がゼロとなることである．

リーマン接続なら接続は計量的で $T = 0$． リーマン接続なら式 (3.11)：

$$\Gamma_{ij,k} = \Gamma^{\ell}{}_{ij}\,g_{\ell k} = \frac{1}{2}\left(\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij}\right)$$

が成り立つ．よって $\Gamma_{ij,k} = \Gamma_{ji,k}$ が成り立ち，$T = 0$ である．また，上式が成り立てば，内積が保存され接続は計量的となる.（既に証明している.）

接続が計量的で $T = 0$ なら接続はリーマン接続となる． 接続が計量的なら

$$\partial_k g_{ij} - \Gamma_{ki,j} - \Gamma_{kj,i} = 0$$

が成り立つ．インデックスを巡回させると

$$\begin{aligned}
\partial_k g_{ij} - \Gamma_{ki,j} - \Gamma_{kj,i} &= 0 \\
\partial_i g_{jk} - \Gamma_{ij,k} - \Gamma_{ik,j} &= 0 \\
\partial_j g_{ki} - \Gamma_{jk,i} - \Gamma_{ji,k} &= 0
\end{aligned}$$

を得る．したがって,

$$\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij} = (\Gamma_{ij,k} + \Gamma_{ik,j}) + (\Gamma_{jk,i} + \Gamma_{ji,k}) - (\Gamma_{ki,j} + \Gamma_{kj,i}) = 2\Gamma_{ij,k}$$

を得る．上式は，リーマン接続となる条件の式に等しい．

3.3　双対アフィン接続

3.3.1　双対アフィン接続の定義

定義アフィン接続 ∇ を持つリーマン多様体 (M, g) において,

$$X g(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X Z) \tag{3.14}$$

で定義されるアフィン接続 ∇^* を，計量 g に関する ∇ の双対アフィン接続という．

この定義から写像 ∇^* は一意に定まる．また，共変微分の公理を満たす．

文献 [2] でも，任意の $Y \in \mathfrak{X}(M)$ に対して式 (3.14) を満たす $\nabla^*_X Z \in \mathfrak{X}(M)$ が一意的に存在することは，内積の一意性（命題 3.4）から明らかとしか言っていない．それほど自明なことなのか？

さらに文献 [2] では以下の証明を行なっている．

• ∇^* は M のアフィン接続を定める．

これを示すには，$X, Y, Z \in \mathfrak{X}(M)$，$f \in C^\infty(M)$ にたいして，∇^* が次の (i)–(iv) を満たすことを示せばよい．

(i) $\nabla^*_{Y+Z} X = \nabla^*_Y X + \nabla^*_Z X$

(ii) $\nabla^*_{fY} X = f \nabla^*_Y X$

(iii) $\nabla^*_Z (X + Y) = \nabla^*_Z X + \nabla^*_Z Y$

(iv) $\nabla^*_Y (fX) = (Yf)X + f \nabla^*_Y X$

証明： (i) 式 (3.14) から，$\forall W \in \mathfrak{X}(M)$ に対して，

$$(Y+Z)g(W, X) = g(\nabla_{Y+Z} W, X) + g(W, \nabla^*_{Y+Z} X)$$

すなわち，

$$g(W, \nabla^*_{Y+Z} X) = (Y+Z)g(W, X) - g(\nabla_{Y+Z} W, X)$$

である．

$$\begin{aligned} g(W, \nabla^*_{Y+Z} X) &= (Y+Z)g(W, X) - g(\nabla_{Y+Z} W, X) \\ &= Yg(W, X) + Zg(W, X) - g(\nabla_Y W + \nabla_Z W, X) = Yg(W, X) + Zg(W, X) - g(\nabla_Y W, X) - g(\nabla_Z W, X) \\ &= g(\nabla_Y W, X) + g(W, \nabla^*_Y X) + g(\nabla_Z W, X) + g(W, \nabla^*_Z, X) - g(\nabla_Y W, X) - g(\nabla_Z W, X) = g(W, (\nabla^*_Y + \nabla^*_Z)X) \end{aligned}$$

すなわち

$$g(W, \nabla^*_{Y+Z} X) = g(W, (\nabla^*_Y + \nabla^*_Z)X)$$

であり，したがって，

$$\nabla^*_{Y+Z} = \nabla^*_Y + \nabla^*_Z$$

が言える．

証明： (ii)

$$fXg(Y, Z) = g(\nabla_{(fX)} Y, Z) + g(Y, \nabla^*_{(fX)} Z)$$

から

$$g(Y, \nabla^*_{(fX)} Z) = fXg(Y, Z) - g(\nabla_{(fX)} Y, Z) = fg(\nabla_X Y, Z) + fg(Y, \nabla^*_X Z) - fg(\nabla_X Y, Z) = g(Y, f\nabla^*_X Z)$$

であり，

$$g(Y, \nabla^*_{(fX)} Z) = g(Y, f\nabla^*_X Z)$$

つまり，

$$\nabla^*_{(fX)} = f\nabla^*_X$$

が言える．

証明： (iii)

$$Zg(W,(X+Y)) = g(\nabla_Z W, X+Y) + g(W, \nabla^*_Z(X+Y))$$

したがって，

$$\begin{aligned} g(W,\nabla^*_Z(X+Y)) &= Zg(W,X+Y) - g(\nabla_Z W, X+Y) = Zg(W,X) + Zg(W,Y) - g(\nabla_Z W, X) - g(\nabla_Z W, Y) \\ &= g(\nabla_Z W, X) + g(W,\nabla^*_Z X) + g(\nabla_Z W, Y) + g(W, \nabla^*_Z Y) - g(\nabla_Z W, X) - g(\nabla_Z W, Y) \\ &= g(W, \nabla^*_Z X + \nabla^*_Z Y) \end{aligned}$$

であるので，

$$\nabla^*_Z(X+Y) = \nabla^*_Z X + \nabla^*_Z Y$$

が言える．

証明： (iv)

$$Yg(W,fX) = g(\nabla_Y W, fX) + g(W, \nabla^*_Y(fX))$$

であり，

$$\begin{aligned} g(W,\nabla^*_Y(fX)) &= Yg(W,fX) - g(\nabla_Y W, fX) = Y(fg(W,fX)) - fg(\nabla_Y W, X) \\ &= (Yf)g(W,X) + fYg(W,X) - fg(\nabla_Y W, X) \\ &= (Yf)g(W,X) + fg(\nabla_Y W, X) + fg(W,\nabla^*_Y X) - fg(\nabla_Y W, X) \\ &= (Yf)g(W,X) + fg(W,\nabla^*_Y X) = g(W, (Yf)X + f\nabla^*_Y X) \end{aligned}$$

がなりたち，

$$\nabla^*_Y(fX) = (Yf)X + f\nabla^*_Y X$$

を得る．**証明終**

∇ が g に対して計量的であることと $\nabla = \nabla^*$ は同値である．

∇ が g に対して計量的であれば，$X, Y, Z \in \mathfrak{X}(M)$ に対して，

$$Xg(Y,Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z) \tag{3.15}$$

が成り立つ．式 (3.14) と比較すれば，

$$\nabla = \nabla^*$$

である．また，上式が成り立てば，式 (3.15) が成り立つので，∇ は計量的であると言える．

定義：　リーマン多様体 (M,g) において，計量 g に関する双対性（式 (3.14)）を満たすアフィン接続のペア (∇, ∇^*) が与えられたとき，3 つの組 (g, ∇, ∇^*) を M の双対構造という．

次が成り立つ．

$$\nabla \text{ がリーマン接続} \iff \nabla = \nabla^* \text{ かつ捩率がゼロ}$$

なぜならば，リーマン接続の定義（接続が計量的）

$$Xg(Y,Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$$

と，式 (3.14) を見比べれば，$\nabla = \nabla^*$ であれば ∇ は計量的である．

逆に，$\nabla \neq \nabla^*$ であれば，一般的には ∇ および ∇^* ともに計量的でない．しかし，

$$\bar{\nabla} = \frac{1}{2}(\nabla + \nabla^*)$$

は常に計量的である．

証明： 双対アフィン接続の定義から，

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X Z)$$

であり，また，Z と Y を入れ替えて，

$$Xg(Z, Y) = g(\nabla_X Z, Y) + g(Z, \nabla^*_X Y)$$

内積の順序を入れ替えて

$$Xg(Z, Y) = g(Z, \nabla^*_X Y) + g(\nabla_X Z, Y) = g(\nabla^*_X Y, Z) + g(Y, \nabla_X Z)$$

一番上の式と足して 2 で割れば，

$$Xg(Z, Y) = g(\bar{\nabla}_X Y, Z) + g(Y, \bar{\nabla}_X Z)$$

を得る．すなわち，$\bar{\nabla}$ は計量的である．

疑問点：双対アフィン接続の特殊な場合（$\nabla = \nabla^*$）がリーマン接続である．すると，双対アフィン接続はリーマン接続よりも広い概念（よりありふれている）のか？

定理：

双対平行移動に関する内積の不変性．曲線 $C = \{p(t); a \leq t \leq b\}$ に沿った ∇ および ∇^* に関する平行移動を Π_C と Π^*_C と書くと，任意の $\boldsymbol{v}, \boldsymbol{w} \in T_{p(a)}M$ に対して，

$$g_{p(b)}(\Pi_C \boldsymbol{v}, \Pi^*_C \boldsymbol{w}) = g_{p(a)}(\boldsymbol{v}, \boldsymbol{w})$$

が成り立つ．つまり，1 つの接ベクトルを ∇ に関して平行移動し，もう 1 つを ∇^* に関して平行移動するなら，内積は保存される．

証明： C に沿って $\nabla-$ 平行なベクトル場を Y，∇^*- 平行なベクトル場を Z とする．

ベクトル場 X を C に沿って微分するとは

$$\nabla_{\dot{p}(t)} X$$

を計算することであり，X が C に沿って平行であるとは，

$$\nabla_{\dot{p}(t)} X = 0$$

が成り立つことである．曲線 $p(t)$ を座標表示すると，$p(t) = (x^1(t), \ldots, x^n(t))$ であり，

$$\dot{p}(t) = \frac{dx^i}{dt} \frac{\partial}{\partial x^i} = \frac{d}{dt}$$

である．さて，今，ベクトル場 Y と Z は C に沿って平行としているので，

$$\nabla_{\dot{p}(t)} Y = 0 \qquad \nabla^*_{\dot{p}(t)} Z = 0$$

である．したがって，双対接続の定義より，双対平行移動に関して

$$\frac{d}{dt}g(Y,Z) = g(\nabla_{\dot{p}(t)}Y, Z) + g(Y, \nabla^*_{\dot{p}(t)}Z) = 0$$

が成り立つ．したがって，$g(Y,Z)$ は定数であり，内積は保存される．

定理： $\nabla-$ 曲率 R がゼロであることと，∇^*- 曲率 R^* がゼロであることは同値である．

証明： 恒等式

$$[X,Y]g(Z,W) = XYg(Z,W) - YXg(Z,W)$$

において，まず左辺は $\nabla-$ 接続と ∇^*- 接続を用いて

$$[X,Y]g(Z,W) = g(\nabla_{[X,Y]}Z, W) + g(Z, \nabla^*_{[X,Y]}W)$$

となる．一方右辺の第 1 項は

$$\begin{aligned} XYg(Z,W) &= X\left[Yg(Z,W)\right] = X\left[g(\nabla_Y Z, W) + g(Z, \nabla^*_Y W)\right] = Xg(\nabla_Y Z, W) + Xg(Z, \nabla^*_Y W) \\ &= g(\nabla_X \nabla_Y Z, W) + g(\nabla_Y Z, \nabla^*_X W) + g(\nabla_X Z, \nabla^*_Y W) + g(Z, \nabla^*_X \nabla^*_Y W) \end{aligned}$$

であり，右辺の第 2 項は

$$\begin{aligned} YXg(Z,W) &= Y\left[Xg(Z,W)\right] = Y\left[g(\nabla_X Z, W) + g(Z, \nabla^*_X W)\right] = Yg(\nabla_X Z, W) + Yg(Z, \nabla^*_X W) \\ &= g(\nabla_Y \nabla_X Z, W) + g(\nabla_X Z, \nabla^*_Y W) + g(\nabla_Y Z, \nabla^*_X W) + g(Z, \nabla^*_Y \nabla^*_X W) \end{aligned}$$

したがって，

$$g(\nabla_{[X,Y]}Z, W) + g(Z, \nabla^*_{[X,Y]}W) = g(\nabla_X \nabla_Y Z, W) + g(Z, \nabla^*_X \nabla^*_Y W) - g(\nabla_Y \nabla_X Z, W) - g(Z, \nabla^*_Y \nabla^*_X W)$$

すなわち，

$$-\left[g(\nabla_X \nabla_Y Z, W) - g(Z, \nabla_Y \nabla_X W) - g(\nabla_{[X,Y]}Z, W)\right] = g(Z, \nabla^*_X \nabla^*_Y W) - g(Z, \nabla^*_Y \nabla^*_X W) - g(Z, \nabla^*_{[X,Y]}W)$$

であるので，

$$-g(R(X,Y,Z)Z, W) = g(Z, R^*(X,Y,Z)W)$$

が成り立つ．上式から，$R=0$ と $R^*=0$ は同値である.（つまり，$R=0$ なら $R^*=0$ であり，$R^*=0$ なら $R=0$ である.）

3.3.2 統計多様体

定理：統計多様体

接続 ∇ が計量的である条件：

$$Xg(Y,Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$$

を思い出し，ここで，

$$(\nabla g)(X,Y,Z) = Xg(Y,Z) - g(\nabla_X Y, Z) - g(Y, \nabla_X Z) \tag{3.16}$$

を定義する．これが，接続 ∇ が「どのくらい計量的でないか」を測るテンソル量である．($(\nabla g)(X, Y, Z)$ は $(\nabla_X g)(Y, Z)$ と書く流儀もある [2]．)

∇ と ∇^* を互いに双対な接続とする．次の 4 条件のうち，任意の 2 条件を仮定すると，残りの 2 条件が成り立つ．

(1) ∇ の捩率はゼロである．

(2) ∇^* の捩率はゼロである．

(3) $(\nabla g)(X, Y, Z)$ は X, Y, Z に関して対称である．(コダッチの方程式)

(4) $\bar{\nabla} = \frac{1}{2}(\nabla + \nabla^*)$ は Riemann 接続である．

証明：　まず，式 (3.16) を書き直す．

$$(\nabla g)(X, Y, Z) = g(\nabla^*_X Y, Z) - g(\nabla_X Y, Z)$$

と書き直せる．（この式の導出は未確認.）X と Y を入れ替えて，

$$(\nabla g)(Y, X, Z) = g(\nabla^*_Y X, Z) - g(\nabla_Y X, Z)$$

であるので，

$$(\nabla g)(X, Y, Z) - (\nabla g)(Y, X, Z) = g(\nabla^*_X Y - \nabla^*_Y X, Z) - g(\nabla_X Y - \nabla_Y X, Z)$$

を得る．また，捩率について，

$$T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y]$$
$$T^*(X, Y) = \nabla^*_X Y - \nabla^*_Y X - [X, Y]$$

であるので，結局，

$$\begin{aligned}(\nabla g)&(X, Y, Z) - (\nabla g)(Y, X, Z) \\ &= g(T^*(X, Y) + [X, Y], Z) - g(T(X, Y) + [X, Y], Z) = g(T^*(X, Y) - T(X, Y), Z)\end{aligned} \tag{3.17}$$

である．

(1),(2) が成り立っている場合：

式 (3.17) より，

$$(\nabla g)(X, Y, Z) = (\nabla g)(Y, X, Z)$$

が成り立つ．Y, Z の対称性は自明であるため，(3) が成り立つ．また，$T = T^* = 0$ が仮定されているため，$\bar{\nabla}$ は Riemann 接続である．したがって，(3),(4) が言える．

(1),(3) を仮定した場合：

式 (3.17) より，$T = T^*$ が言えるので，$T^* = 0$ が言える．上と同じ論法で $\bar{\nabla}$ は Riemann 接続である．(2),(3) を仮定した場合も同じである．

(1),(4) を仮定した場合：

$\nabla^* = 2\bar{\nabla} - \nabla$ を

$$T^*(X, Y) = \nabla^*_X Y - \nabla^*_Y X - [X, Y]$$

に代入すれば,

$$T^*(X,Y) = (2\bar{\nabla}_X - \nabla_X)Y - (2\bar{\nabla}_Y - \nabla_Y)X - [X,Y] = 2\bar{T}(X,Y) - T(X,Y) = 0$$

なぜなら，$\bar{\nabla}$ は Riemann 接続なので，$\bar{T} = 0$．したがって，(2) が言える．(1),(2) から (3) が言える．(2),(4) を仮定した場合も同じである．

(3),(4) を仮定した場合：

式 (3.17) より,

$$g(T^*(X,Y) - T(X,Y), Z) = 0$$

Z は任意なので,

$$T^*(X,Y) - T(X,Y) = 0$$

(4) より，$\bar{\nabla}$ はリーマン接続なので，その捩率

$$\frac{1}{2}(T(X,Y) + T^*(X,Y)) = 0 \quad \rightarrow \quad T(X,Y) + T^*(X,Y) = 0$$

したがって，$T = T^* = 0$ が言える．

文献 [2] における (1),(2)→(3) の証明： まず最初に以下の成立を示す.

$$(\nabla g)(X,Y,Z) + g(Y, T(X,Z)) = (\nabla g)(Z,Y,X) + g(Y, T^*(X,Z))$$

（文献 [2] では，$(\nabla g)(X,Y,Z)$, $(\nabla g)(Z,Y,X)$ はそれぞれ $(\nabla_X g)(Y,Z)$, $(\nabla_Z g)(Y,X)$ と表されている.）

証明：

$$\begin{aligned}(\nabla g)(X,Y,Z) + g(Y,T(X,Z)) &= Xg(Y,Z) - g(\nabla_X Y, Z) - g(Y, \nabla_X Z) + g(Y, \nabla_X Z - \nabla_Z X - [X,Z]) \\ &= Xg(Y,Z) - g(\nabla_X Y, Z) - g(Y, \nabla_X Z) + g(Y, \nabla_X Z) - g(Y, \nabla_Z X) - g(Y, [X,Z]) \\ &= Xg(Y,Z) - g(\nabla_X Y, Z) - g(Y, \nabla_Z X) - g(Y, [X,Z])\end{aligned}$$

また,

$$\begin{aligned}(\nabla g)(Z,Y,X) + g(Y,T^*(X,Z)) &= Zg(Y,X) - g(\nabla_Z Y, X) - g(Y, \nabla_Z X) + g(Y, \nabla^*_X Z - \nabla^*_Z X - [X,Z]) \\ &= Zg(Y,X) - g(\nabla_Z Y, X) - g(Y, \nabla_Z X) + g(Y, \nabla^*_X Z) - g(Y, \nabla^*_Z X) - g(Y, [X,Z])\end{aligned}$$

である．ここで，双対アフィン接続の定義式から

$$g(Y, \nabla^*_X Z) = Xg(Y,Z) - g(\nabla_X Y, Z)$$

$$g(Y, \nabla^*_Z X) = Zg(Y,X) - g(\nabla_Z Y, X)$$

であるので，代入して,

$$\begin{aligned}&(\nabla g)(Z,Y,X) + g(Y,T^*(X,Z)) \\ &= Zg(Y,X) - g(\nabla_Z Y, X) - g(Y, \nabla_Z X) + Xg(Y,Z) - g(\nabla_X Y, Z) - Zg(Y,X) + g(\nabla_Z Y, X) - g(Y, [X,Z]) \\ &= Xg(Y,Z) - g(\nabla_X Y, Z) - g(Y, \nabla_Z X) - g(Y, [X,Z])\end{aligned}$$

であり，したがって，

$$(\nabla g)(X,Y,Z) + g(Y, T(X,Z)) = (\nabla g)(Z,Y,X) + g(Y, T^*(X,Z))$$

が成り立つ．したがって，$T = 0, T^* = 0$ が成り立てば，

$$(\nabla g)(X,Y,Z) = (\nabla g)(Z,Y,X)$$

がなりたつ．上記を（あるいは前述の条件 (3) を）コダッチの方程式と呼ぶ．

定義：統計多様体

多様体 M にアフィン接続 ∇ とリーマン計量 g が与えられているとき，∇ の捩率がゼロでコダッチの方程式が成り立つとき (M, ∇, g) を統計多様体という．

これまでの議論から，上記は，以下の定義と同じである．

多様体 M にアフィン接続 ∇ とリーマン計量 g が与えられているとき，∇ とその相対接続 ∇^* の捩率がともにゼロであるとき (M, ∇, g) を統計多様体という．

あるいは，前述の条件 (1)–(4) が満たされている多様体を統計多様体と呼ぶ．(ただし，4 つの条件のうち 2 つが満たされていればよい.）

3.3.3　双対平坦な多様体

定義：　双対構造 (g, ∇, ∇^*) を持つ多様体 M において，∇ に関する曲率 R も捩率 T もともにゼロとなり，かつ ∇^* に関する曲率 R と捩率 T がゼロとなるとき，M **は双対平坦**であるという．

平坦な多様体上では局所アフィン座標系（接続が恒等的にゼロとなる座標系）が存在する．接続 ∇ と ∇^* について同時に平坦な多様体を考えているので，局所アフィン座標系もそれぞれの接続ごとに 2 種類存在する．そこで，$U; x^1, \dots, x^n)$ を ∇ に関するアフィン座標近傍，$V; y^1, \dots, y^n)$ を ∇^* に関するアフィン座標近傍とする．

定理：

双対平坦な多様体 M では，各点の周りで

$$g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^j}\right) = \delta_{ij}$$

を満たす局所 $\nabla-$ アフィン座標系 (x^i) と，局所 ∇^*- アフィン座標系 (y^j) を取ることができる．この特性を持った (x^i) と (y^j) を**双対アフィン座標系**と呼ぶ．

証明：　点 $p \in M$ の近傍において ∇ に関するアフィン座標近傍 $(U; \xi^i)$ と ∇^* に関するアフィン座標近傍 $(V; \eta^i)$ とを定める．そして，

$$g_p\left(\left(\frac{\partial}{\partial \xi^i}\right)_p, \left(\frac{\partial}{\partial \eta^j}\right)_p\right)$$

を (i,j) 成分とする行列を $\boldsymbol{G}$ とする．計量の正定値性より，$\boldsymbol{G}$ は正則行列である．ここで，列ベクトル $\boldsymbol{\xi} = [\xi^1, \dots, \xi^n]^T$，$\boldsymbol{\eta} = [\eta^1, \dots, \eta^n]^T$ として，

$$\boldsymbol{x} = \boldsymbol{\xi}$$
$$\boldsymbol{y} = \boldsymbol{G}\boldsymbol{\eta}$$

とする新しい座標系を導入する．新しい座標系は $\boldsymbol{x} = [x^1, \ldots, x^n]^T$，$\boldsymbol{y} = [y^1, \ldots, y^n]^T$ であらわした．$\boldsymbol{y} = \boldsymbol{G\eta}$ を成分で表すと，

$$y^j = \sum_k G_{jk}\eta^k \quad \to \quad \eta^k = \sum_j [G^{-1}]_{kj} y^j$$

と表される．したがって，

$$\frac{\partial}{\partial y^j} = \sum_k \frac{\partial \eta^k}{\partial y^j}\frac{\partial}{\partial \eta^k} = \sum_k [G^{-1}]_{kj}\eta^k$$

である．ここで，

$$g\left(\left(\frac{\partial}{\partial x^i}\right)_p, \left(\frac{\partial}{\partial y^j}\right)_p\right) g\left(\left(\frac{\partial}{\partial x^i}\right)_p, \left(\sum_k [G^{-1}]_{kj}\frac{\partial}{\partial \eta^k}\right)_p\right)$$
$$= \sum_k [G^{-1}]_{kj} g\left(\left(\frac{\partial}{\partial x^i}\right)_p, \left(\frac{\partial}{\partial \eta^k}\right)_p\right) = \sum_k [G^{-1}]_{kj} G_{ik} = \delta_{ij}$$

である．したがって，上で定義した (x^i) と (y^j) が求める座標系である．

以下は必要か？　任意の p で上が証明できていればそれで OK なのではないのか．

こうして作ったアフィン座標系の組 $(x^i), (y^j)$ がすべての $p \in U \cap V$ で

$$g\left(\left(\frac{\partial}{\partial x^i}\right)_p, \left(\frac{\partial}{\partial y^j}\right)_p\right) = \delta_{ij}$$

を満たすことを示す．これを示すには，任意のベクトル場 $X \in \mathfrak{X}(M)$ に対して，

$$Xg\left(\left(\frac{\partial}{\partial x^i}\right)_p, \left(\frac{\partial}{\partial y^j}\right)_p\right) = 0$$

が言えれば，$g\left(\left(\frac{\partial}{\partial x^i}\right)_p, \left(\frac{\partial}{\partial y^j}\right)_p\right) = \delta_{ij}$ がどこでも成り立つことが言える．実際，

$$Xg\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^j}\right) = g\left(\nabla_X \frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^j}\right) + g\left(\frac{\partial}{\partial x^i}, \nabla^*_X \frac{\partial}{\partial y^j}\right)$$

であるが，(x^i) が ∇ アフィン座標系なので

$$\nabla_X \frac{\partial}{\partial x^i} = 0$$

(y^i) が ∇^* アフィン座標系なので

$$\nabla^*_X \frac{\partial}{\partial y^i} = 0$$

である[6]．したがって，$Xg = 0$ であり，$g\left(\left(\frac{\partial}{\partial x^i}\right)_p, \left(\frac{\partial}{\partial y^j}\right)_p\right) = \delta_{ij}$ が $p \in U \cap V$ のどこでも成り立つ．

[6]補足：

$$\nabla_X \frac{\partial}{\partial x^i} = \nabla_{X^j \frac{\partial}{\partial x^j}} \frac{\partial}{\partial x^i} = X^j \Gamma^k{}_{ji} \frac{\partial}{\partial x^k}$$

であるが，アフィン座標系なので $\Gamma^k{}_{ji}$ は恒等的にゼロであり，$\nabla_X \frac{\partial}{\partial x^i} = 0$ である．$\nabla^*_X \frac{\partial}{\partial y^i} = 0$ についても全く同様な議論が成り立つ．

記号の定義 以下のように記号を定義する．上で議論した直交性を満たす $\nabla-$ アフィン座標系を (θ^i) で，∇^*- アフィン座標系を η_j で表す．これらを $\theta-$ 座標系，$\eta-$ 座標系と呼ぶ．

対応する接ベクトル場の基底を

$$\partial_i = \frac{\partial}{\partial\theta^i} \qquad \partial^i = \frac{\partial}{\partial\eta_i}$$

と書く.（添え字が上付きか下付きかで θ^i の接ベクトルか η_i の接ベクトルかを区別する．わかりにくい!!　要注意）そして，直交性は

$$g(\partial_i, \partial^j) = g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\eta_j}\right) = \delta_i^j$$

と表す.

3.3.4　双対座標とルジャンドル変換

この節は，このように定義した $\theta-$ 座標系と $\eta-$ 座標系がルジャンドル変換の関係にあることを証明するものである.

まず，準備として以下の補題を証明する.

補題：双対アフィン座標系 $\{\theta^i, \eta_i\}$ に対する計量 g の成分を

$$g_{ij} = g(\partial_i, \partial_j) = g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\theta^j}\right) \qquad g^{ij} = g(\partial^i, \partial^j) = g\left(\frac{\partial}{\partial\eta_i}, \frac{\partial}{\partial\eta_j}\right)$$

とおくと，

$$g_{ij} = \partial_i\eta_j = \frac{\partial\eta_j}{\partial\theta^i} = \frac{\partial\eta_i}{\partial\theta^j} = \partial_j\eta_i$$

$$g^{ij} = \partial^i\theta^j = \frac{\partial\theta^j}{\partial\eta_i} = \frac{\partial\theta^i}{\partial\eta_j} = \partial^j\theta^i$$

$$g_{ij}g^{jk} = \delta_i^k$$

が成り立つ.

証明：

$$\partial_i = \frac{\partial}{\partial\theta^i} = \frac{\partial\eta_k}{\partial\theta^i}\frac{\partial}{\partial\eta_k} = \frac{\partial\eta_k}{\partial\theta^i}\partial^k$$

であるので，

$$g_{ij} = g(\partial_i, \partial_j) = g(\frac{\partial\eta_k}{\partial\theta^i}\partial^k, \partial_j) = \frac{\partial\eta_k}{\partial\theta^i}g(\partial^k, \partial_j) = \frac{\partial\eta_k}{\partial\theta^i}\delta_j^k = \frac{\partial\eta_j}{\partial\theta^i} = \partial_i\eta_j \tag{3.18}$$

が成り立つ．ここで，g_{ij} は座標変換 $(\theta^i) \to (\eta_j)$ の座標変換のヤコビ行列 $\left(\frac{\partial\eta_j}{\partial\theta^i}\right)$ である．また，計量は対称であるので（内積は順序によらないので）

$$g_{ij} = g_{ji} = \frac{\partial\eta_i}{\partial\theta^j} = \partial_j\eta_i$$

も成り立つ.

また，

$$\partial^i = \frac{\partial}{\partial\eta_i} = \frac{\partial\theta^k}{\partial\eta_i}\frac{\partial}{\partial\theta^k} = \frac{\partial\theta^k}{\partial\eta_i}\partial^k$$

したがって，

$$g^{ij} = g(\partial^i, \partial^j) = \frac{\partial\theta^k}{\partial\eta_i} g(\partial^k, \partial_j) = \frac{\partial\theta^k}{\partial\eta_i}\delta^k_j = \frac{\partial\theta^j}{\partial\eta_i} = \partial^i\theta^j \tag{3.19}$$

である．したがって，g^{ij} は座標変換 $(\eta_i) \to (\theta^j)$ の座標変換のヤコビ行列 $\left(\dfrac{\partial\theta^j}{\partial\eta_i}\right)$ である．また，計量の対称性から

$$g^{ij} = g^{ji} = \partial^j\theta^i = \frac{\partial\theta^i}{\partial\eta_j}$$

も成り立つ．最後に

$$g_{ij}g^{jk} = \frac{\partial\eta_j}{\partial\theta^i}\frac{\partial\theta^k}{\partial\eta_j} = \delta^k_i$$

は，ヤコビ行列とその逆行列を乗じているので単位行列になる．

定理：　(M, g, ∇, ∇^*) を n 次元双対平坦空間，$\theta = (\theta^1, \ldots, \theta^n)$，$\eta = (\eta_1, \ldots, \eta_n)$ を M の局所双対アフィン座標系とする．g_{ij} と g^{ij} は

$$g_{ij} = \left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\theta^j}\right) \qquad g^{ij} = \left(\frac{\partial}{\partial\eta_i}, \frac{\partial}{\partial\eta_j}\right)$$

と定める．すると，ある関数 $\psi(\theta)$ と $\varphi(\eta)$ が存在し，

$$g_{ij} = \frac{\partial^2\psi}{\partial\theta^i\partial\theta^j} \qquad g_{ij} = \frac{\partial^2\varphi}{\partial\eta_i\partial\eta_j}$$

と表され，また，

$$\theta^i = \frac{\partial\varphi}{\partial\eta_i} \quad \eta_i = \frac{\partial\psi}{\partial\theta^i} \quad \psi(\theta) + \varphi(\eta) = \theta^k\eta_k$$

が成り立つ．ここで，$\psi(\theta)$，$\varphi(\eta)$ はポテンシャルと呼ばれる[7]．つまり，θ^i はポテンシャル $\varphi(\eta)$ の η_i に関する勾配として，η^j はポテンシャル $\psi(\theta)$ の θ_j に関する勾配として求まるのである．

証明：　関係式：

$$\frac{\partial\eta_i}{\partial\theta^j} = \frac{\partial\eta_j}{\partial\theta^i}$$

は連立微分方程式

$$\eta_i = \frac{\partial\psi}{\partial\theta^i} \qquad i = 1, \ldots, n$$

において，ポテンシャル関数 ψ が解として存在することを意味する．（これは，微分方程式における解の存在条件に関する議論であり，このノートではこの議論に係わらず，結論をそのまま信じる．）したがって，式 (3.18) を用いて，

$$g_{ij} = \frac{\partial\eta_i}{\partial\theta^j} = \frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}$$

を得る．また，関係式：

$$\frac{\partial\theta^j}{\partial\eta_i} = \frac{\partial\theta^i}{\partial\eta_j}$$

は連立微分方程式

$$\theta^i = \frac{\partial\varphi}{\partial\eta_i} \qquad i = 1, \ldots, n$$

[7] θ と η はそれぞれ $\theta^1, \ldots, \theta^n$ と $\eta_1, \ldots, \eta_n$ を集合的に表したもので，$\psi(\theta) = \psi(\theta^1, \ldots, \theta^n)$，$\varphi(\eta) = \varphi(\eta_1, \ldots, \eta_n)$ の意味である．

において，ポテンシャル関数 φ が解として存在することを意味する．したがって，式 (3.19) を用いて

$$g^{ij} = \frac{\partial \theta^i}{\partial \eta_j} = \frac{\partial^2 \varphi}{\partial \eta_i \partial \eta_j}$$

を得る．

$\psi + \varphi - \theta^k \eta_k$ の全微分を計算すると

$$d\psi + d\varphi - (d\theta^k)\eta_k - \theta^k(d\eta_k) = \frac{\partial \psi}{\partial \theta^i} d\theta^i + \frac{\partial \varphi}{\partial \eta_i} d\eta_i - \eta_i d\theta^i - \theta^i d\eta_i = \eta_i d\theta^i + \theta^i d\eta_i - \eta_i d\theta^i - \theta^i d\eta_i = 0$$

したがって，$\psi + \varphi - \theta^k \eta_k$ は定数関数である．ポテンシャル関数には定数倍の任意性があるので，ポテンシャル関数に定数を加えることにより

$$\psi + \varphi - \theta^k \eta_k = 0 \tag{3.20}$$

とできる．

さらに，ポテンシャル関数は $p, q \in M$ に対して

$$\varphi(\eta(p)) = \max_{q \in M} \left(\theta^i(q)\eta_i(p) - \psi(\theta(q)) \right) \tag{3.21}$$

$$\psi(\theta(p)) = \max_{q \in M} \left(\eta_i(q)\theta^i(p) - \varphi(\eta(q)) \right) \tag{3.22}$$

と表される．

証明 [式 (3.21)**]：**p は固定し，q で微分すれば，

$$\begin{aligned}\frac{d}{dq}\left(\theta^i(q)\eta_i(p) - \psi(\theta(q))\right) &= \left(\frac{d\theta^i(q)}{dq}\eta_i(p) - \frac{d\psi(\theta(q))}{dq} \right) \\ &= \left(\eta_i(p) - \frac{\partial \psi(\theta(q))}{\partial \theta^i(q)} \right) \frac{d\theta^i(q)}{dq} = (\eta_i(p) - \eta_i(q)) \frac{d\theta^i(q)}{dq}\end{aligned}$$

である．したがって，式 (3.21) の max はすべての i で $\eta_i(p) = \eta_i(q)$，すなわち，$p = q$ のときに達成される．その最大値は

$$\varphi(\eta(p)) = \theta^i(p)\eta_i(p) - \psi(\theta(p))$$

を満たす．これは式 (3.20) が成り立つ条件である．式 (3.22) も同様に証明できる．

3.4　ダイバージェンス

3.4.1　定義：$\nabla-$ ダイバージェンス

(M, g, ∇, ∇^*) を双対平坦多様体とする．

2 点 $p, q \in M$ に対して決まる量

$$D(p\|q) = \psi(\theta(p)) + \varphi(\eta(q)) - \theta^i(p)\eta_i(q)$$

を $\nabla-$ ダイバージェンスと呼ぶ．ここで，θ^i, η_i は M の双対アフィン座標系である．

以下に示すように，$\nabla-$ ダイバージェンスは双対アフィン座標系の取り方にはよらない．

3.4.2 $\nabla-$ ダイバージェンス：座標変換不変性

定理：

θ^i, η_i と $\widetilde{\theta}^i$ と $\widetilde{\eta}_i$ を M の2組の双対アフィン座標系とする．それぞれの双対ポテンシャルを $\psi(\theta), \varphi(\eta)$，$\widetilde{\psi}(\widetilde{\theta}), \widetilde{\varphi}(\widetilde{\eta})$ と書く．すると，

$$\psi(\theta(p)) + \varphi(\eta(q)) - \theta^i(p)\eta_i(q) = \widetilde{\psi}(\widetilde{\theta}(p)) + \widetilde{\varphi}(\widetilde{\eta}(q)) - \widetilde{\theta}^i(p)\widetilde{\eta}_i(q)$$

が成り立つ．

証明：

アフィン座標系はあるアフィン変換で関係付けられている．

$$\widetilde{\theta}^\lambda = A_i^\lambda \theta^i + a^\lambda \qquad \widetilde{\eta}_\lambda = B_\lambda^i \eta_i + b_\lambda \tag{3.23}$$

$\boldsymbol{A}$ と $\boldsymbol{B}$ は正則行列である．$\widetilde{\partial}_\lambda = \dfrac{\partial}{\partial \widetilde{\theta}^\lambda}$ と書けば，ここで，

$$\frac{\partial}{\partial \theta^i} = \frac{\partial \widetilde{\theta}^\lambda}{\partial \theta^i}\frac{\partial}{\partial \widetilde{\theta}^\lambda} \quad \rightarrow \quad \partial_i = \frac{\partial \widetilde{\theta}^\lambda}{\partial \theta^i}\widetilde{\partial}_\lambda = A_i^\lambda \widetilde{\partial}_\lambda$$

であり，同様に

$$\frac{\partial}{\partial \eta_i} = \frac{\partial \widetilde{\eta}_\lambda}{\partial \eta_i}\frac{\partial}{\partial \widetilde{\eta}_\lambda} \quad \rightarrow \quad \partial^i = \frac{\partial \widetilde{\eta}_\lambda}{\partial \eta_i}\widetilde{\partial}^\lambda = B_\lambda^i \widetilde{\partial}^\lambda$$

である．

双対アフィン座標系においては $g(\partial_i, \partial^j) = g(\dfrac{\partial}{\partial \theta^i}, \dfrac{\partial}{\partial \eta_j}) = \delta_i^j$ の関係がある．したがって，

$$\delta_i^j = g(\partial_i, \partial^j) = g(A_i^\lambda \widetilde{\partial}_\lambda, B_\mu^j \widetilde{\partial}^\mu) = A_i^\lambda B_\mu^j g(\widetilde{\partial}_\lambda, \widetilde{\partial}^\mu) = A_i^\lambda B_\mu^j g(\widetilde{\partial}_\lambda, \widetilde{\partial}^\mu)\delta_\lambda^\mu = A_i^\lambda B_\lambda^j$$

を得る．ここで，行列 $\boldsymbol{A}$ と $\boldsymbol{B}$ を

$$\boldsymbol{A} = \begin{bmatrix} A_1^1 & \cdots & A_1^n \\ \vdots & \ddots & \vdots \\ A_n^1 & \cdots & A_n^n \end{bmatrix} \qquad \boldsymbol{B} = \begin{bmatrix} B_1^1 & \cdots & B_1^n \\ \vdots & \ddots & \vdots \\ B_n^1 & \cdots & B_n^n \end{bmatrix}$$

とすれば，すなわち，上付きの添え字が列を，下付きが行を表すように行列を定義すれば

$$A_i^\lambda B_\lambda^j = \delta_i^j$$

は

$$\boldsymbol{AB} = \boldsymbol{I}$$

を表す．先ほど導いた式

$$\partial_i = A_i^\lambda \widetilde{\partial}_\lambda \quad \rightarrow \quad \begin{bmatrix} \partial_1 \\ \vdots \\ \partial_n \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} \widetilde{\partial}_1 \\ \vdots \\ \widetilde{\partial}_n \end{bmatrix} \quad \rightarrow \quad \begin{bmatrix} \widetilde{\partial}_1 \\ \vdots \\ \widetilde{\partial}_n \end{bmatrix} = \boldsymbol{B} \begin{bmatrix} \partial_1 \\ \vdots \\ \partial_n \end{bmatrix} \quad \rightarrow \quad \widetilde{\partial}_\lambda = B_\lambda^i \partial_i$$

全く同様に

$$\partial^i = B^i_\lambda \widetilde{\partial}^\lambda \quad \to \quad [\partial_1, \ldots, \partial_n] = \boldsymbol{B}[\widetilde{\partial}_1, \ldots, \widetilde{\partial}_n] \quad \to \quad [\widetilde{\partial}_1, \ldots, \widetilde{\partial}_n] = \boldsymbol{A}[\partial_1, \ldots, \partial_n \quad \to \quad \widetilde{\partial}^\lambda = A^\lambda_i \partial^i$$

も成り立つ．

いよいよ証明に入る．まず，ψ と $\widetilde{\psi}$ の関係を導く．$\eta_i = \dfrac{\partial \psi}{\partial \theta^i}$ から，

$$\widetilde{\eta}_\lambda = \frac{\partial \widetilde{\psi}}{\partial \widetilde{\theta}^\lambda} = \frac{\partial}{\partial \widetilde{\theta}^\lambda} \widetilde{\psi} = B^j_\lambda \frac{\partial}{\partial \theta^j} \widetilde{\psi}$$

一方，

$$\widetilde{\eta}_\lambda = B^j_\lambda \eta_j + b_\lambda = B^j_\lambda \frac{\partial \psi}{\partial \theta^j} + b_\lambda$$

上 2 つの式を見比べて，

$$B^j_\lambda \frac{\partial}{\partial \theta^j} \widetilde{\psi} = B^j_\lambda \frac{\partial}{\partial \theta^j} \psi + b_\lambda$$

$$\frac{\partial}{\partial \theta^j} \widetilde{\psi} = \frac{\partial}{\partial \theta^j} \psi + [\boldsymbol{B}^{-1}]^\lambda_j b_\lambda \quad \to \quad \frac{\partial}{\partial \theta^j} \widetilde{\psi} = \frac{\partial}{\partial \theta^j} \psi + A^\lambda_j b_\lambda$$

を得る．両辺を θ^j で積分して，

$$\widetilde{\psi}(\widetilde{\theta}) = \psi(\theta) + A^\lambda_j b_\lambda \theta^j + C \tag{3.24}$$

を得る．ここで，C は積分定数である，

次に，φ と $\widetilde{\varphi}$ の関係を導く．

$$\psi(\theta) + \varphi(\eta) = \theta^k \eta_k$$

を用いれば，

$$\begin{aligned}\widetilde{\varphi}(\widetilde{\eta}) = \widetilde{\theta}^k \widetilde{\eta}_k - \widetilde{\psi}(\widetilde{\theta}) &= (A^k_i \theta^i + a^k)(B^j_k \eta_j + b_k) - (\psi(\theta) + A^\lambda_j b_\lambda \theta^j + C) \\ &= \theta^i \eta_i - \psi(\theta) + a^k B^j_k \eta_j + a^k b_k - C = \varphi(\eta) + a^k B^j_k \eta_j + a^k b_k - C \end{aligned} \tag{3.25}$$

を得る．したがって，式 (3.23), (3.24), (3.25) を用いれば，

$$\begin{aligned}&\widetilde{\psi}(\widetilde{\theta}(p)) + \widetilde{\varphi}(\widetilde{\eta}(q)) - \widetilde{\theta}^k(p)\widetilde{\eta}_k(q) \\ &\quad = \big(\psi(\theta(p)) + A^\lambda_j b_\lambda \theta^j(p) + C\big) + \Big(\varphi(\eta(q)) + a^k B^j_k \eta_j(q) + a^k b_k - C\Big) - (A^k_i \theta^i(p) + a^k)(B^j_k \eta_j(q) + b_k) \\ &\quad = \psi(\theta(p)) + \varphi(\eta(q)) - \theta^i(p)\eta_i(q)\end{aligned} \tag{3.26}$$

を得る．すなわち，ダイバージェンスは座標変換に対して不変であり，幾何学的に固有な量である．

3.4.3　ダイバージェンスの性質

$p \in M$ の θ 座標と η 座標を

$$\theta(p) = (\theta^1(p), \ldots, \theta^n(p)) \qquad \eta(p) = (\eta_1(p), \ldots, \eta_n(p))$$

と表せば，式 (3.21) と (3.22)：

$$\varphi(\eta(p)) = \max_{q\in M}\big(\theta^i(q)\eta_i(p) - \psi(\theta(q))\big)$$
$$\psi(\theta(p)) = \max_{q\in M}\big(\eta_i(q)\theta^i(p) - \varphi(\eta(q))\big)$$

が成り立つ．この事から，$p \neq q$ の場合では，

$$\varphi(\eta(p)) \geq \theta^i(q)\eta_i(p) - \psi(\theta(q)) \to \varphi(\eta(p)) + \psi(\theta(q)) - \theta^i(q)\eta_i(p) \geq 0$$

および

$$\psi(\theta(q)) \geq \eta_i(p)\theta^i(q) - \varphi(\eta(p)) \to \psi(\theta(q)) + \varphi(\eta(p)) - \eta_i(p)\theta^i(q) \geq 0$$

であり，

$$D(p\|q) \geq 0$$

を得る．また，等号成立は $p = q$ の場合のみである．

3.4.4 ユークリッド空間での例

$\nabla^* = \nabla$ である場合は自己双対と呼ばれる．これは，リーマン接続をもつリーマン多様体の場合である．この場合，双対平坦性は，単なる平坦性であり，平坦なリーマン多様体はユークリッド空間に帰着する．

$$\psi(\theta) + \varphi(\eta) = \theta^k\eta_k$$

において，$\theta^i = \eta_i = z^i$ として，$\psi(z) = \varphi(z)$ とすれば，

$$2\psi(z) = \sum_{i=1}^n (z^i)^2$$

すなわち，

$$\psi(z) = \varphi(z) = \frac{1}{2}\sum_{i=1}^n (z^i)^2$$

を得る．したがって，

$$D(p\|q) = \psi(z(p)) + \psi(z(q)) - \sum_{i=1}^n z^i(p)z^i(q) = \frac{1}{2}\sum_{i=1}^n (z^i(p))^2 + \frac{1}{2}\sum_{i=1}^n (z^i(q))^2 - \sum_{i=1}^n z^i(p)z^i(q) = \frac{1}{2}\sum_{i=1}^n (z^i(p) - z^i(q))^2$$

となる．

ダイバージェンスは距離の 2 乗の次元を持っているが，ユークリッド距離とは異なる（$\frac{1}{2}$ がついている.）．さらに，距離の公理のうちで，対称性や三角不等式はなりたたない．

3.4.5 一般化したピタゴラスの定理

一般化したピタゴラスの定理が成り立つ．

双対平坦多様体 (M, g, ∇, ∇^*) 上に 3 点 p, q, r を取る．もし，p と q を結ぶ $\nabla-$ 測地線が，q と r を結ぶ ∇^*- 測地線が q において直交するなら

$$D(p\|q) + D(q\|r) = D(p\|r)$$

が成り立つ．証明省略

3.4.6　指数分布族の例：KL ダイバージェンス

パラメータ $\boldsymbol{\theta}$ を持つ確率分布が指数分布族であるとは，確率分布が

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = h(\boldsymbol{x})g(\boldsymbol{\theta})\exp[\theta^T \boldsymbol{u}(\boldsymbol{x})]$$

と表される場合を言う (Bishop, PRML,p110)．ここで，$\boldsymbol{\theta} = (\theta^1, \dots, \theta^n)$ であり，$\boldsymbol{u}(\boldsymbol{x})$ は任意のベクトル関数 $\boldsymbol{u}(\boldsymbol{x}) = (u_1(\boldsymbol{x}), \dots, \boldsymbol{u}_n(\boldsymbol{x}))$ である．この確率分布は

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \exp[c(\boldsymbol{x}) + \theta^i u_i(\boldsymbol{x}) - \psi(\boldsymbol{\theta})] \tag{3.27}$$

と表しておく．ここで，$h(\boldsymbol{x}) = \exp(c(\boldsymbol{x}))$，$g(\boldsymbol{\theta}) = \exp(-\psi(\boldsymbol{\theta}))$ である[8].

ここで，$\boldsymbol{\theta}$ の双対アフィン局所座標系を求める．**と言うことは，指数分布族で表される確率分布多様体は双対平坦性を持っているのか？**

式 (3.27) の両辺を θ^i で微分すると，

$$\begin{aligned}\frac{\partial}{\partial\theta^i}p(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{\partial}{\partial\theta^i}\exp[c(\boldsymbol{x}) + \theta^i u_i(\boldsymbol{x})] - \psi(\boldsymbol{\theta})] &= \left(u_i(\boldsymbol{x}) - \frac{\partial\psi}{\partial\theta^i}(\boldsymbol{\theta})\right)\exp[c(\boldsymbol{x}) + \theta^i u_i(\boldsymbol{x})] - \psi(\boldsymbol{\theta})] \\ &= \left(u_i(\boldsymbol{x}) - \frac{\partial\psi}{\partial\theta^i}(\boldsymbol{\theta})\right)p(\boldsymbol{x}|\boldsymbol{\theta})\end{aligned} \tag{3.28}$$

を得る．ここで，Ω を確率変数 $\boldsymbol{x}$ の値域として，

$$\sum_{\boldsymbol{x}\in\Omega} p(\boldsymbol{x}|\boldsymbol{\theta}) = 1$$

に注意すれば，

$$\frac{\partial}{\partial\theta^i}\sum_{\boldsymbol{x}\in\Omega} p(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\Omega}\left(u_i(\boldsymbol{x}) - \frac{\partial\psi}{\partial\theta^i}(\boldsymbol{\theta})\right)p(\boldsymbol{x}|\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\left[u_i(\boldsymbol{x}) - \frac{\partial\psi}{\partial\theta^i}(\boldsymbol{\theta})\right] = E_{\boldsymbol{\theta}}\left[u_i(\boldsymbol{x})\right] - \frac{\partial\psi}{\partial\theta^i}(\boldsymbol{\theta}) = 0 \tag{3.29}$$

である．上式で，$E_{\boldsymbol{\theta}}$ は確率分布関数 $p(\boldsymbol{x}|\boldsymbol{\theta})$ による期待値を表す．ここで，$E_{\boldsymbol{\theta}}\left[\frac{\partial\psi}{\partial\theta^i}(\boldsymbol{\theta})\right] = \frac{\partial\psi}{\partial\theta^i}(\boldsymbol{\theta})$ であることに注意.（なぜなら，この項は確率変数 $\boldsymbol{x}$ を含まないので.）さらに，

$$\log p(\boldsymbol{x}|\boldsymbol{\theta}) = c(\boldsymbol{x}) + \theta^i u_i(\boldsymbol{x}) - \psi(\boldsymbol{\theta}) \tag{3.30}$$

であるので，

$$\frac{\partial^2}{\partial\theta^i\partial\theta^j}\log p(\boldsymbol{x}|\boldsymbol{\theta}) = -\frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}(\boldsymbol{\theta}) \tag{3.31}$$

---

[8]例えば，正規分布

$$p(x|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

の場合，

$$\log p(x|\boldsymbol{\theta}) = \log\left[\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]\right] = -\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x - \left(\frac{\mu^2}{2\sigma^2} + \log(\sqrt{2\pi}\sigma)\right)$$

であるので，

$$c(x) = 0, \quad \theta^1 = -\frac{1}{2\sigma^2}, \quad u_1(x) = x^2, \quad \theta^2 = \frac{\mu}{\sigma^2}, \quad u_2(x) = x$$

であり，

$$\psi(\theta^1, \theta^2) = -\frac{(\theta^2)^2}{4\theta^1} + \frac{1}{2}\log\left(-\frac{\pi}{\theta^1}\right)$$

である．

が成り立つ．

ところで，フィシャー計量の定義式は

$$g_{ij} = \sum_{\boldsymbol{x}\in\Omega} \left(\frac{\partial}{\partial\theta^i}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta^j}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)p(\boldsymbol{x}|\boldsymbol{\theta})$$

である．ここで，

$$\left(\frac{\partial}{\partial\theta^j}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)p(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{1}{p(\boldsymbol{x}|\boldsymbol{\theta})}p(\boldsymbol{x}|\boldsymbol{\theta})\left(\frac{\partial}{\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta})\right) = \frac{\partial}{\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta})$$

を用いれば，

$$g_{ij} = \sum_{\boldsymbol{x}\in\Omega}\left(\frac{\partial}{\partial\theta^i}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)\frac{\partial}{\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta})$$

となる．ここで，

$$g_{ij} = \sum_{\boldsymbol{x}\in\Omega}\left(\frac{\partial}{\partial\theta^i}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)\frac{\partial}{\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta}) = -\sum_{\boldsymbol{x}\in\Omega}\left(\frac{\partial^2}{\partial\theta^i\partial\theta^j}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)p(\boldsymbol{x}|\boldsymbol{\theta}) \tag{3.32}$$

を示すことができる．上式 (3.32) は（若干長くなるため）本節の終わりに証明する．式 (3.31) を用いると，

$$g_{ij} = -\sum_{\boldsymbol{x}\in\Omega}\left(\frac{\partial^2}{\partial\theta^i\partial\theta^j}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)p(\boldsymbol{x}|\boldsymbol{\theta}) = -\sum_{\boldsymbol{x}\in\Omega}\left(-\frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}(\boldsymbol{\theta}))\right)p(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}(\boldsymbol{\theta}) \tag{3.33}$$

を得る．

ここで，

$$\eta_i = E_{\boldsymbol{\theta}}\left[u_i(\boldsymbol{x})\right]$$

とすれば，式 (3.29) から，

$$\eta_i = \frac{\partial\psi}{\partial\theta^i}(\boldsymbol{\theta})$$

を得，また，式 (3.33) から

$$g_{ij} = \frac{\partial\eta_i}{\partial\theta^j} = \frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}$$

であるので，θ^i と η_i は互いに双対なアフィン局所座標系となっている．

次に，$\psi(\boldsymbol{\theta}) + \varphi(\boldsymbol{\eta}) = \theta^i\eta_i$ を満たす，ルジャンドル変換のポテンシャル $\varphi(\boldsymbol{\eta})$ を求める．まず，

$$\log p(\boldsymbol{x},|\theta) = c(\boldsymbol{x}) + \theta^i u_i(\boldsymbol{x}) - \psi(\boldsymbol{\theta})$$

であるので，両辺の期待値を取ると，

$$E_{\boldsymbol{\theta}}[\log p(\boldsymbol{x},|\theta) - c(\boldsymbol{x})] = E_{\boldsymbol{\theta}}[\theta^i u_i(\boldsymbol{x}) - \psi(\boldsymbol{\theta})] = \theta^i E_{\boldsymbol{\theta}}[u_i(\boldsymbol{x})] - \psi(\boldsymbol{\theta}) = \theta^i\eta_i - \psi(\boldsymbol{\theta})$$

したがって，

$$\varphi(\boldsymbol{\eta}) = \theta^i\eta_i - \psi(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}[\log p(\boldsymbol{x},|\theta) - c(\boldsymbol{x})]$$

である．上式は $p, q \in M$ で成り立つので

$$\varphi(\boldsymbol{\eta}(p)) = E_{\boldsymbol{\theta}(p)}[\log p(\boldsymbol{x},|\theta) - c(\boldsymbol{x})]$$
$$\varphi(\boldsymbol{\eta}(q)) = E_{\boldsymbol{\theta}(q)}[\log q(\boldsymbol{x},|\theta) - c(\boldsymbol{x})]$$

である．

M 上に 2 点 p, q を取り，ダイバージェンスを計算すれば

$$D(p\|q) = \psi(\boldsymbol{\theta}(p)) + \varphi(\boldsymbol{\eta}(q)) - \theta^i(p)\eta_i(q) = \psi(\boldsymbol{\theta}(p)) + E_{\boldsymbol{\theta}(q)}[\log q(\boldsymbol{x}, |\theta) - c(\boldsymbol{x})] - \theta^i E_{\boldsymbol{\theta}(q)}[u_i(\boldsymbol{x})]$$

となる．ここで，

$$\psi(\boldsymbol{\theta}(p)) - \theta^i(p)E_{\boldsymbol{\theta}(q)}[u_i(\boldsymbol{x})] = -E_{\boldsymbol{\theta}(q)}[\log p(\boldsymbol{x}|\boldsymbol{\theta}) - c(\boldsymbol{x})]$$

であるのでダイバージェンス $D(p\|q)$ は

$$\begin{aligned} D(p\|q) &= E_{\boldsymbol{\theta}(q)}[\log q(\boldsymbol{x}, |\theta) - c(\boldsymbol{x})] - E_{\boldsymbol{\theta}(q)}[\log p(\boldsymbol{x}|\boldsymbol{\theta}) - c(\boldsymbol{x})] \\ &= E_{\boldsymbol{\theta}(q)}[\log q(\boldsymbol{x}, |\theta) - c(\boldsymbol{x}) - \log p(\boldsymbol{x}|\boldsymbol{\theta}) + c(\boldsymbol{x})] = E_{\boldsymbol{\theta}(q)}[\log q(\boldsymbol{x}, |\theta) - \log p(\boldsymbol{x}|\boldsymbol{\theta})] \\ &= \sum_{\boldsymbol{x}\in\Omega} q(\boldsymbol{x}|\boldsymbol{\theta}) \log \frac{q(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x}|\boldsymbol{\theta})} \end{aligned} \tag{3.34}$$

となる．これは KL ダイバージェンスに等しい．

疑問点のまとめ

1) 双対平坦な多様体では,「距離」は定義できないのか？「出来ない」とすればダイバージェンスは距離に代わる概念なのか？

２）距離が定義できるなら，距離以外にダイバージェンスを定義することににどのような意味が有るのか？

式 (3.32) の証明

$$g_{ij} = \sum_{\boldsymbol{x}\in\Omega} \left(\frac{\partial}{\partial\theta^i} \log p(\boldsymbol{x}|\boldsymbol{\theta}) \right) \frac{\partial}{\partial\theta^j} p(\boldsymbol{x}|\boldsymbol{\theta}) = -\sum_{\boldsymbol{x}\in\Omega} \left(\frac{\partial^2}{\partial\theta^i\partial\theta^j} \log p(\boldsymbol{x}|\boldsymbol{\theta}) \right) p(\boldsymbol{x}|\boldsymbol{\theta})$$

を示す．逆からスタートすれば

$$\begin{aligned} &\sum_{\boldsymbol{x}\in\Omega} \left(\frac{\partial^2}{\partial\theta^i\partial\theta^j} \log p(\boldsymbol{x}|\boldsymbol{\theta}) \right) p(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\Omega} \left[\frac{\partial}{\partial\theta^i} \left(\frac{\partial}{\partial\theta^j} \log p(\boldsymbol{x}|\boldsymbol{\theta}) \right) \right] p(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\Omega} \left[\frac{\partial}{\partial\theta^i} \frac{\frac{\partial}{\partial\theta^j} p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x}|\boldsymbol{\theta})} \right] p(\boldsymbol{x}|\boldsymbol{\theta}) \\ &= \sum_{\boldsymbol{x}\in\Omega} \left[\frac{\left[\frac{\partial^2}{\partial\theta^i\partial\theta^j} p(\boldsymbol{x}|\boldsymbol{\theta}) \right] p(\boldsymbol{x}|\boldsymbol{\theta}) - \left[\frac{\partial}{\partial\theta^i} p(\boldsymbol{x}|\boldsymbol{\theta}) \right] \left[\frac{\partial}{\partial\theta^j} p(\boldsymbol{x}|\boldsymbol{\theta}) \right]}{p(\boldsymbol{x}|\boldsymbol{\theta})^2} \right] p(\boldsymbol{x}|\boldsymbol{\theta}) \\ &= \sum_{\boldsymbol{x}\in\Omega} \frac{\partial^2}{\partial\theta^i\partial\theta^j} p(\boldsymbol{x}|\boldsymbol{\theta}) - \sum_{\boldsymbol{x}\in\Omega} \left[\frac{\partial}{\partial\theta^i} p(\boldsymbol{x}|\boldsymbol{\theta}) \right] \left[\frac{\partial}{\partial\theta^j} p(\boldsymbol{x}|\boldsymbol{\theta}) \right] \frac{1}{p(\boldsymbol{x}|\boldsymbol{\theta})} \end{aligned}$$

上式の右辺第 1 項は,

$$\sum_{\boldsymbol{x}\in\Omega} \frac{\partial^2}{\partial\theta^i\partial\theta^j} p(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{\partial^2}{\partial\theta^i\partial\theta^j} \sum_{\boldsymbol{x}\in\Omega} p(\boldsymbol{x}|\boldsymbol{\theta}) = 0$$

である．また，第 2 項は（先ほどのトリックを使って）

$$\begin{aligned}\sum_{\boldsymbol{x}\in\Omega}\left[\frac{\partial}{\partial\theta^i}p(\boldsymbol{x}|\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta})\right]\frac{1}{p(\boldsymbol{x}|\boldsymbol{\theta})} &= \sum_{\boldsymbol{x}\in\Omega}\left[\frac{1}{p(\boldsymbol{x}|\boldsymbol{\theta})}\frac{\partial}{\partial\theta^i}p(\boldsymbol{x}|\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta})\right] \\ &= \sum_{\boldsymbol{x}\in\Omega}\left[\frac{\partial}{\partial\theta^i}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta})\right]\end{aligned}$$

であるので，

$$\sum_{\boldsymbol{x}\in\Omega}\left(\frac{\partial^2}{\partial\theta^i\partial\theta^j}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right)p(\boldsymbol{x}|\boldsymbol{\theta}) = -\sum_{\boldsymbol{x}\in\Omega}\left[\frac{\partial}{\partial\theta^i}\log p(\boldsymbol{x}|\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta^j}p(\boldsymbol{x}|\boldsymbol{\theta})\right]$$

を，すなわち，式 (3.32) を示すことができた．

第4章　指数型分布族の多様体と双対平坦性

2023-6-12

前ノートにおいて，簡単な確率分布空間の例としてスカラー正規分布多様体について考察した．その結果，この2次元多様体は定曲率空間ではあるものの，平坦な空間ではなく，２つの正規分布間の距離なども多様体の曲率を考慮しなければならないことを議論した．

正規分布空間は，フィシャー計量のもとでリーマン接続を考える限り平坦とはならないが，リーマン接続の条件（１）アフィン接続,（２）捩率ゼロ,（３）接続が計量的，のうちで（３）の条件をあきらめることで，正規分布多様体に対して「双対平坦性」という性質をもつ接続（双対接続）を導出できる．

双対平坦性や双対接続は一般的な概念であり，正規分布空間に固有な概念ではないが，正規分布を含む指数型分布族と呼ばれる確率分布の空間では特に有用である．このノートでは，指数型分布族に議論を限定して双対平坦性を解説する．

4.1　曲率と平坦性

4.1.1　確率分布族の多様体

$p(x|\boldsymbol{\theta})$ を確率変数 x でパラメータ $\boldsymbol{\theta} = (\theta^1, \ldots, \theta^n)$ を持つ確率（密度）分布とする．確率（密度）分布は以下の条件を満たす．

$$\{p(x|\boldsymbol{\theta}) : \Omega \to \mathbb{R} \,|\, p(\boldsymbol{x}|\boldsymbol{\theta}) > 0, \int_\Omega p(\boldsymbol{x}|\boldsymbol{\theta})dx = 1\}$$

すると，パラメータ $\boldsymbol{\theta}$ を持った確率分布を座標 $\boldsymbol{\theta} = (\theta^1, \ldots, \theta^n)$ における「点」とみなした n 次元多様体を考えることができる．この多様体を

$$M = \{p_{\boldsymbol{\theta}} = p(\boldsymbol{x}|\boldsymbol{\theta}) \,|\, \boldsymbol{\theta} = (\theta^1, \ldots, \theta^n) \in \Theta\}$$

と表し，統計モデルとよぶ．つまり，統計モデルとは確率分布を（パラメータの値を座標とする）点と考える多様体のことである（「統計モデルは確率分布を点と考える多様体と同一視される」と言う言い方もある）．ただし，この場合，パラメータ領域 Θ がユークリッド空間の開集合と同相で，確率分布が θ^i で微分可能で，フィシャー情報行列が多様体のリーマン計量となることが必要である.（確率分布の（ある種の）不変性からこれが必要となる.）

4.1.2　アフィン接続

定義　抽象的な曲面（多様体）M 上において，以下の4つの条件を満たす写像

$$\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M) \quad : \quad (\boldsymbol{X}, \boldsymbol{Y}) \mapsto \nabla_{\boldsymbol{X}}\boldsymbol{Y}$$

を M 上の共変微分と呼ぶ.

M 上のベクトル場を $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \in \mathfrak{X}(M)$ とし，任意の微分可能な（C^∞ 級）関数を f とする.

(i) $\nabla_{\boldsymbol{X}+\boldsymbol{Y}} \boldsymbol{Z} = \nabla_{\boldsymbol{X}} \boldsymbol{Z} + \nabla_{\boldsymbol{Y}} \boldsymbol{Z}$

(ii) $\nabla_{f\boldsymbol{X}} \boldsymbol{Z} = f \nabla_{\boldsymbol{X}} \boldsymbol{Z}$

(iii) $\nabla_{\boldsymbol{X}} (\boldsymbol{Y} + \boldsymbol{Z}) = \nabla_{\boldsymbol{X}} \boldsymbol{Y} + \nabla_{\boldsymbol{X}} \boldsymbol{Z}$

(iv) $\nabla_{\boldsymbol{X}} (f\boldsymbol{Y}) = (\boldsymbol{X} f) \nabla_{\boldsymbol{X}} \boldsymbol{Y} + f \nabla_{\boldsymbol{X}} \boldsymbol{Y}$

ここで，ベクトル場 $\boldsymbol{X}$ と $\boldsymbol{Y}$ に対して，写像

$$(\boldsymbol{X}, \boldsymbol{Y}) \quad \mapsto \quad \nabla_{\boldsymbol{X}} \boldsymbol{Y}$$

が上記 (i) から (iv) を満たせば,

$$\nabla_{\frac{\partial}{\partial u^i}} \frac{\partial}{\partial u^j} = {\Gamma^k}_{ij} \frac{\partial}{\partial u^k} \tag{4.1}$$

なる ${\Gamma^k}_{ij}$ を定めれば $\nabla_{\boldsymbol{X}} \boldsymbol{Y}$ が計算できる[1].

多様体 M の局所座標系に式 (4.1) を満たす微分可能な関数の組 ${\Gamma^k}_{ij}$ を与えることを，M にアフィン接続を与えると言い，${\Gamma^k}_{ij}$ を接続係数とよぶ．アフィン接続を与えることと，共変微分（すなわち上の定義による写像）を与えることは等価であり，通常,「共変微分」と「アフィン接続」は同じ意味を持つと考える（同一視する）.

アフィン接続の条件 (i)–(iv) は，結局，接続に対して（多重）線形性が成り立つ条件である.

4.1.3 捩率

$$T(\boldsymbol{X}, \boldsymbol{Y}) = \nabla_{\boldsymbol{X}} \boldsymbol{Y} - \nabla_{\boldsymbol{Y}} \boldsymbol{X} - [\boldsymbol{X}, \boldsymbol{Y}]$$

はベクトル場を 2 個取って 1 個のベクトル場を返す写像,

$$T : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$$

---

[1]実際，M 上のベクトル場の基底は

$$\frac{\partial}{\partial u^i} \quad i = 1, \ldots, n$$

で表される．したがって，M 上のベクトル場 $\boldsymbol{X}, \boldsymbol{Y}$ を

$$\boldsymbol{X} = X^i \frac{\partial}{\partial u^i} \qquad \boldsymbol{Y} = Y^j \frac{\partial}{\partial u^j}$$

と表すことができる．すると，上に述べた性質 (i)–(iv) を用いれば,

$$\nabla_{\boldsymbol{X}} \boldsymbol{Y} = \nabla_{X^i \frac{\partial}{\partial u^i}} \boldsymbol{Y} = X^i \nabla_{\frac{\partial}{\partial u^i}} \boldsymbol{Y} = X^i \nabla_{\frac{\partial}{\partial u^i}} \left(Y^j \frac{\partial}{\partial u^j} \right) = X^i \left(\frac{\partial Y^j}{\partial u^i} \frac{\partial}{\partial u^j} + Y^j \nabla_{\frac{\partial}{\partial u^i}} \frac{\partial}{\partial u^j} \right)$$

となる．ここで，式 (4.1) にしたがって ${\Gamma^k}_{ij}$ が定められているとすれば,

$$\nabla_{\boldsymbol{X}} \boldsymbol{Y} = X^i \left(\frac{\partial Y^j}{\partial u^i} \frac{\partial}{\partial u^j} + Y^j \nabla_{\frac{\partial}{\partial u^i}} \frac{\partial}{\partial u^j} \right) = X^i \left(\frac{\partial Y^j}{\partial u^i} \frac{\partial}{\partial u^j} + Y^j {\Gamma^k}_{ij} \frac{\partial}{\partial u^k} \right) = \left(X^i \frac{\partial Y^k}{\partial u^i} + X^i Y^j {\Gamma^k}_{ij} \right) \frac{\partial}{\partial u^k}$$

を得る.

すなわち，$(0,2)$ 型のテンソルであり，捩率と呼ばれる．接続係数の満たす条件に，捩率がゼロ，

$$T(\boldsymbol{X},\boldsymbol{Y})=0$$

を追加する．

$$\nabla_{\boldsymbol{X}}\boldsymbol{Y}=\left(X^i\frac{\partial Y^k}{\partial u^i}+X^iY^j\Gamma^k{}_{ij}\right)\frac{\partial}{\partial u^k}$$

$$\nabla_{\boldsymbol{Y}}\boldsymbol{X}=\left(Y^j\frac{\partial Y^k}{\partial u^j}+X^iY^j\Gamma^k{}_{ji}\right)\frac{\partial}{\partial u^k}$$

であるので，

$$\nabla_{\boldsymbol{X}}\boldsymbol{Y}-\nabla_{\boldsymbol{Y}}\boldsymbol{X}=\left(X^i\frac{\partial Y^k}{\partial u^i}-Y^j\frac{\partial Y^k}{\partial u^j}\right)\frac{\partial}{\partial u^k}+X^iY^j\left(\Gamma^k{}_{ij}-\Gamma^k{}_{ji}\right)\frac{\partial}{\partial u^k}$$

ここで，

$$\left(X^i\frac{\partial Y^k}{\partial u^i}-Y^j\frac{\partial Y^k}{\partial u^j}\right)\frac{\partial}{\partial u^k}=[\boldsymbol{X},\boldsymbol{Y}]$$

であるので，結局，

$$\nabla_{\boldsymbol{X}}\boldsymbol{Y}-\nabla_{\boldsymbol{Y}}\boldsymbol{X}=[\boldsymbol{X},\boldsymbol{Y}]+X^iY^j\left(\Gamma^k{}_{ij}-\Gamma^k{}_{ji}\right)\frac{\partial}{\partial u^k}$$

である．すなわち捩率がゼロが成り立つ条件は，アフィン接続係数が対称性

$$\Gamma^k{}_{ij}=\Gamma^k{}_{ji}$$

を持つことと等価である．なお，捩率については，手短な説明が [3] にある．

4.1.4 リーマン（レビ・チヴィタ）接続

リーマン接続は，(1) アフィン接続，(2) 捩率ゼロ，の条件に，次の条件

$$\boldsymbol{X}(\boldsymbol{Y},\boldsymbol{Z})=(\nabla_{\boldsymbol{X}}\boldsymbol{Y},\boldsymbol{Z})+(\boldsymbol{Y},\nabla_{\boldsymbol{X}}\boldsymbol{Z}) \tag{4.2}$$

を付け加えて得られる接続である．式 (4.2) は内積に対して，微分のライプニッツ則が成り立つことを要請するものであり，この条件のもとで，接続係数は 1 つに決まり，局所座標 $(x^1,\ldots,x^n)$ を用いて，

$$\Gamma^\ell{}_{ij}=g^{\ell k}\left(\frac{\partial g_{jk}}{\partial x^i}+\frac{\partial g_{ki}}{\partial x^j}-\frac{\partial g_{ij}}{\partial x^k}\right) \tag{4.3}$$

として求まる．上式の接続はリーマン接続，あるいはレビ・チヴィタ接続とよばれ微分幾何学で普通に使われる接続である．また，式 (4.3) の接続は内積を保存することが知られており，この性質を持つ接続は計量的と呼ばれる．

リーマン（レビ・チヴィタ）接続-補足

リーマン接続の式 (4.3)：

$$\Gamma_{ij,k}=\Gamma^\ell{}_{ij}\,g_{\ell k}=\frac{1}{2}\left(\partial_i g_{jk}+\partial_j g_{ki}-\partial_k g_{ij}\right)$$

は座標系に依存しない形で表すと

$$g(\nabla_X Y,Z)=\frac{1}{2}\left(Xg(Y,Z)+Yg(Z,X)-Zg(X,Y)-g(X,[Y,Z])+g(Y,[Z,X])+g(Z,[X,Y])\right) \tag{4.4}$$

と表される．計量性の条件

$$\boldsymbol{X} g(\boldsymbol{Y}, \boldsymbol{Z}) = g(\nabla_X \boldsymbol{Y}, \boldsymbol{Z}) + g(\boldsymbol{Y}, \nabla_X \boldsymbol{Z})$$

と捩率がゼロであること

$$T = \nabla_X Y - \nabla_Y X - [X, Y] = 0$$

から，式 eqrefRISaa が成り立つ（左辺と右辺は等しいこと）は簡単に証明できる．さらに，式 (4.3) に等しいことは以下に示す．

局所座標を用いて

$$\boldsymbol{X} = \partial_k \qquad \boldsymbol{Y} = \partial_i \qquad \boldsymbol{Z} = \partial_j$$

とすれば，

$$\nabla_X \boldsymbol{Y} = \nabla_{\partial_k} \partial_i = \Gamma^{\ell}{}_{ki}\, \partial_\ell$$

である．したがって，

$$g(\nabla_X Y, Z) = \Gamma^{\ell}{}_{ki}\, g(\partial_\ell, \partial_j) = \Gamma^{\ell}{}_{ki}\, g_{\ell j} = \Gamma_{ki,j}$$

である．また，

$$X g(Y, Z) = \partial_k g_{ij} \qquad Y g(Z, X) = \partial_i g_{jk} \qquad Z g(X, Y) = \partial_j g_{ki}$$

である．さらに，

$$[Y, Z] = [\partial_i, \partial_j] = 0 \qquad [Z, X] = [\partial_j, \partial_k] = 0 \qquad [X, Y] = [\partial_k, \partial_i] = 0$$

であるので，式 (4.4) は

$$\Gamma_{ki,j} = \frac{1}{2}\left(\partial_k g_{ij} + \partial_i g_{jk} - \partial_j g_{ki}\right)$$

と導出でき，ここで，添え字を i, j, k の順番で 1 ステップまわすと，

$$\Gamma_{ij,k} = \frac{1}{2}\left(\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij}\right)$$

であり，リーマン接続の式 (4.3) に等しくなる．

4.1.5 曲率テンソルの定義

次に，曲率テンソルについて概説する．

多様体の上でベクトル場 $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \in \mathfrak{X}(M)$ に対して，$\nabla_X \nabla_Y \boldsymbol{Z}$ と $\nabla_Y \nabla_X \boldsymbol{Z}$ は同じにはならない．すなわち，$\nabla_X \nabla_Y$ と $\nabla_Y \nabla_X$ は交換可能ではない．この差が曲面の曲がり方に関係する．

ここで，3 つのベクトル場の入力に対して，ベクトル場を 1 つ返す写像，

$$R : \mathfrak{X}(M) \times \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$$

を考え，

$$R(\boldsymbol{X}, \boldsymbol{Y})\boldsymbol{Z} = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X \boldsymbol{Z} - \nabla_{[X,Y]} \boldsymbol{Z}$$

と定義し，曲率テンソル場と呼ぶ．これは $(1, 3)$ 型（1 階反変，3 階共変）テンソルである．

4.1.6 曲率テンソル場の性質

上で定義した曲率テンソルについて，以下が成り立つ．ただし，$\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}\in\mathfrak{X}(M)$，$f\in C^\infty(M)$（$f$ はいくらでも微分可能）である．2)，3) は R がテンソル（多重線形写像）であることの帰結である．

1) $R(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Z}=-R(\boldsymbol{Y},\boldsymbol{X})\boldsymbol{Z}$

2) $R(f\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Z}=fR(\boldsymbol{Y},\boldsymbol{X})\boldsymbol{Z},\quad R(\boldsymbol{X},f\boldsymbol{Y})\boldsymbol{Z}=fR(\boldsymbol{Y},\boldsymbol{X})\boldsymbol{Z}$

3) $R(\boldsymbol{X},\boldsymbol{Y})(f\boldsymbol{Z})=fR(\boldsymbol{Y},\boldsymbol{X})\boldsymbol{Z}$

さらに，任意のベクトル場 $\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}\in\mathfrak{X}(M)$ に対して，ビアンキの恒等式と呼ばれる

$$R(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Z}+R(\boldsymbol{Y},\boldsymbol{Z})\boldsymbol{X}+R(\boldsymbol{Z},\boldsymbol{X})\boldsymbol{Y}=0$$

が成り立つ．

これらの証明は [1]，[2] にあるが，前ノート「双対アフィン接続とダイバージェンス」にもある．

4.1.7 曲率テンソルの成分表示

局所座標系を用いて曲率テンソル $R(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Z}$ を成分で表す．曲率テンソルは，基底表示をすると

$$R(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{Z}=R^{\ell}_{ijk}dx^k\otimes dx^j\otimes dx^i\otimes\frac{\partial}{\partial x^\ell}$$

であり，テンソルの局所座標系 (x^i) における成分 R^{ℓ}_{ijk} は

$$R(\frac{\partial}{\partial u^k},\frac{\partial}{\partial u^j})\frac{\partial}{\partial u^i}=R^{\ell}_{ijk}\frac{\partial}{\partial u^\ell}$$

として求まる．上式左辺を計算する．

まず，

$$R(\frac{\partial}{\partial u^k},\frac{\partial}{\partial u^j})\frac{\partial}{\partial u^i}=\nabla_{\frac{\partial}{\partial u^k}}\nabla_{\frac{\partial}{\partial u^j}}\frac{\partial}{\partial u^i}-\nabla_{\frac{\partial}{\partial u^j}}\nabla_{\frac{\partial}{\partial u^k}}\frac{\partial}{\partial u^i}-\nabla_{[\frac{\partial}{\partial u^k},\frac{\partial}{\partial u^j}]}\frac{\partial}{\partial u^i}$$

を計算する[2]．ここで，$[\frac{\partial}{\partial u^k},\frac{\partial}{\partial u^j}]=0$ であるので，

$$R(\frac{\partial}{\partial u^k},\frac{\partial}{\partial u^j})\frac{\partial}{\partial u^i}=\nabla_{\frac{\partial}{\partial u^k}}\nabla_{\frac{\partial}{\partial u^j}}\frac{\partial}{\partial u^i}-\nabla_{\frac{\partial}{\partial u^j}}\nabla_{\frac{\partial}{\partial u^k}}\frac{\partial}{\partial u^i}$$

となって，曲率テンソル $R(\frac{\partial}{\partial u^k},\frac{\partial}{\partial u^j})\frac{\partial}{\partial u^i}$ は共変微分の順序を変えた差に等しくなる．

ここで，

$$\nabla_{\frac{\partial}{\partial u^i}}\frac{\partial}{\partial u^j}=\sum_{\ell}\Gamma^{\ell}_{ij}\frac{\partial}{\partial u^\ell}=\Gamma^{\ell}_{ij}\frac{\partial}{\partial u^\ell}$$

[2] R^{ℓ}_{ijk} に関して，教科書にある公式を導くには微分の順序に注意しなければならない．すなわち微分の順序を $\boldsymbol{p}_{ijk}$ と同じにしなければならない．つまり，

$$\boldsymbol{p}_{ijk}=\frac{\partial}{\partial u^k}\left(\frac{\partial}{\partial u^j}\left(\frac{\partial}{\partial u^i}\boldsymbol{p}\right)\right)$$

であるので，$i\to j\to k$ の順番で微分をするように添え字を決める．

を用いる．この式は，接空間の j 方向への基底 $\dfrac{\partial}{\partial u^j}$ の i 方向への共変微分を，${\Gamma^k}_{ij}$ を係数として基底 $\dfrac{\partial}{\partial u^k}$ で展開したものである．すると，

$$\begin{aligned}\nabla_{\frac{\partial}{\partial u^k}}(\nabla_{\frac{\partial}{\partial u^j}}\frac{\partial}{\partial u^i}) = \nabla_{\frac{\partial}{\partial u^k}}({\Gamma^\ell}_{ji}\frac{\partial}{\partial u^\ell}) &= \frac{\partial {\Gamma^\ell}_{ji}}{\partial u^k}\frac{\partial}{\partial u^\ell} + {\Gamma^\ell}_{ji}\nabla_{\frac{\partial}{\partial u^k}}\frac{\partial}{\partial u^\ell} \\ &= \frac{\partial {\Gamma^\ell}_{ji}}{\partial u^k}\frac{\partial}{\partial u^\ell} + {\Gamma^\ell}_{ji}{\Gamma^m}_{k\ell}\frac{\partial}{\partial u^m} = \frac{\partial {\Gamma^\ell}_{ij}}{\partial u^k}\frac{\partial}{\partial u^\ell} + {\Gamma^m}_{ij}{\Gamma^\ell}_{km}\frac{\partial}{\partial u^\ell}\end{aligned}$$

また，

$$\nabla_{\frac{\partial}{\partial u^j}}\nabla_{\frac{\partial}{\partial u^k}}\frac{\partial}{\partial u^i} = \frac{\partial {\Gamma^\ell}_{ik}}{\partial u^j}\frac{\partial}{\partial u^\ell} + {\Gamma^m}_{ik}{\Gamma^\ell}_{jm}\frac{\partial}{\partial u^\ell}$$

したがって，

$$\begin{aligned}R(\frac{\partial}{\partial u^k},\frac{\partial}{\partial u^j})\frac{\partial}{\partial u^i} &= \frac{\partial {\Gamma^\ell}_{ji}}{\partial u^k}\frac{\partial}{\partial u^\ell} + {\Gamma^m}_{ji}{\Gamma^\ell}_{im}\frac{\partial}{\partial u^\ell} - \left(\frac{\partial {\Gamma^\ell}_{ik}}{\partial u^j}\frac{\partial}{\partial u^\ell} + {\Gamma^m}_{ik}{\Gamma^\ell}_{jm}\frac{\partial}{\partial u^\ell}\right) \\ &= \left(\frac{\partial {\Gamma^\ell}_{ji}}{\partial u^k} - \frac{\partial {\Gamma^\ell}_{ik}}{\partial u^j} + {\Gamma^m}_{ji}{\Gamma^\ell}_{km} - {\Gamma^m}_{ik}{\Gamma^\ell}_{jm}\right)\frac{\partial}{\partial u^\ell} \\ &= \left(\frac{\partial {\Gamma^n}_{ij}}{\partial u^k} - \frac{\partial {\Gamma^n}_{ik}}{\partial u^j} + {\Gamma^m}_{ij}{\Gamma^n}_{km} - {\Gamma^m}_{ik}{\Gamma^n}_{jm}\right)\frac{\partial}{\partial u^n} \qquad (4.5)\end{aligned}$$

を得る．これを，

$$R(\frac{\partial}{\partial u^k},\frac{\partial}{\partial u^j})\frac{\partial}{\partial u^i} = R^\ell_{ijk}\frac{\partial}{\partial u^\ell}$$

とする．したがって，

$$R^\ell_{ijk} = \frac{\partial {\Gamma^\ell}_{ij}}{\partial u^k} - \frac{\partial {\Gamma^\ell}_{ik}}{\partial u^j} + {\Gamma^m}_{ij}{\Gamma^\ell}_{km} - {\Gamma^m}_{ik}{\Gamma^\ell}_{jm} \qquad (4.6)$$

を得る.（注意!　いくつかの教科書では，共変微分の順序を変えた場合のちがいから式 (4.6) を導いている．この場合，右辺の項の正負が上式と逆になっている.）

4.1.8　平坦な多様体

平坦性の定義　アフィン接続を持つ多様体 M において，曲率テンソル場 R も捩率テンソル場 T もともに恒等的にゼロとなるとき，M は平坦であると言う.

アフィン座標系：定義　アフィン接続 ∇ を持つ多様体 M の局所座標系 $(x^1, x^2, \ldots, x^n)$ において，接続係数 ${\Gamma^k}_{ij}$ が全て恒等的にゼロとなるとき，この座標系をアフィン座標系と呼ぶ.

例： 3 次元ユークリッド空間は（当たり前であるが）平坦な空間である．この空間に定義された直交座標系では接続係数 ${\Gamma^k}_{ij}$ は全て恒等的にゼロであり，直交座標系はアフィン座標系である．しかしながら，例えば，3 次元ユークリッド空間に定義された極座標系ではノンゼロの接続係数を持つのでアフィン座標系ではない.

定理　アフィン接続 ∇ を持つ多様体 M において，次の 2 条件は同値である.

(i) M は平坦である.

(ii) M にアフィン座標系が存在する.

証明：　(ii)→(i)

この部分は簡単である．アフィン座標系においては接続係数 $\Gamma^k{}_{ij}$ は全てゼロであるので，このとき，曲率テンソル R も捩率テンソル T もゼロとなる．

証明： (i)→(ii)

（これが成立しない場合，つまり接続係数 $\Gamma^k{}_{ij}$ は全てゼロという座標系が存在しない場合には，その空間は平坦では無いわけである．）

証明は Appendix の 4.7.1 節に記載．

さらに以下の定理が成り立つ．

定理 M に局所座標系 $(x^1,\ldots,x^n)$ と $(\xi^1,\ldots,\xi^n)$ があって，(x^i) はアフィン座標系であったとする．このとき，(ξ^a) もアフィン座標系であるための必要十分条件は座標変換がアフィン変換（線形変換）で書けること，すなわち，

$$\begin{bmatrix} x^1 \\ \vdots \\ x^n \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} \xi^1 \\ \vdots \\ \xi^n \end{bmatrix} + \boldsymbol{b}$$

と書けることである．

証明： 証明は簡単である．接続係数の座標変換は

$$\Gamma^k{}_{ij} = \frac{\partial^2 \xi^c}{\partial x^i \partial x^j}\frac{\partial x^k}{\partial \xi^c} + \frac{\partial \xi^b}{\partial x^j}\frac{\partial \xi^a}{\partial x^i}\frac{\partial x^k}{\partial \xi^c}\Gamma^c{}_{ab}$$

と表される．$\Gamma^k{}_{ij} = 0$ のもとで $\Gamma^c{}_{ab} = 0$ となるための必要十分条件は

$$\frac{\partial^2 \xi^c}{\partial x^i \partial x^j} = 0$$

となることであり，これは，座標変換がアフィン変換（座標変換が線形変換 $\xi^c = ax^i + b$）であることと同値である．

3 次元ユークリッド空間の例を用いれば，アフィン座標系である直交座標系に対して，線形変換で得られる新しい座標系はアフィン座標系である.（直交座標系から線形変換で得られるこの新しい座標系もまた直交座標系であるので，当たり前である.）

4.2 双対接続

4.2.1 リーマン接続と双対アフィン接続

リーマン接続に関する議論を復習すると，リーマン接続は，(1) アフィン接続，(2) 捩率ゼロ，の条件に，次の条件（方向微分の内積に対するライプニッツ則：(4.2)）ベクトル場 X, Y, Z に対して，

$$Xg(Y,Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X \boldsymbol{Z})$$

を付け加えて得られる接続である．この条件のもとで，接続係数は 1 つに決まり，局所座標 $(x^1,\ldots,x^n)$ を用いて，

$$\Gamma^\ell{}_{ij} = g^{\ell k}\left(\frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ki}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k}\right) \tag{4.7}$$

として求まる．また，式 (4.3) の計量は内積を保存することが知られておりこの性質を持つ接続は計量的と呼ばれる．

計量的な接続（リーマン接続）は，微分幾何学の多くの分野で普通に用いられているが，スカラー正規分布に対する計算例でもわかるように確率分布族の多様体においてリーマン接続を用いると，多様体はゼロでない曲率を持ってしまう．そこで,（情報幾何では）リーマン接続を用いるのをあきらめ，以下に述べる双対アフィン接続を用いる．双対アフィン接続を用いることで一般的な確率分布の空間を「平坦」とみなせるのである．ただし，カッコ付きの平坦性であり双対平坦と呼ばれる.（と書いたが，情報幾何でなぜ双対接続にこだわるのか，そこまで判然としない．平坦性だけでもないように思う.）

まず，双対アフィン接続を定義する．アフィン接続 ∇ を持つリーマン多様体 (M, g) において，

$$Xg(Y,Z) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X Z) \tag{4.8}$$

を満たすベクトル場 $\nabla^*_X Z \in \mathfrak{X}(M)$ が一意に存在する.（∇ に対して，(4.8) を満たす ∇^* が一意的に存在することはどのように示すことができるか？）さらに，∇^* はアフィン接続である．∇^* を，計量 g に関する ∇ の双対アフィン接続という．∇^* がアフィン接続の性質を満たすことを以下に示す.

4.2.2　双対アフィン接続 ∇^* の持つ性質

(I) 双対アフィン接続 ∇^* は，$X, Y, Z \in \mathfrak{X}(M)$，$f \in C^\infty(M)$ にたいして，次のアフィン接続の性質 (i)–(iv) を満たす.（これが ∇^* が双対「アフィン」接続と呼ばれる理由である.）

(i) $\nabla^*_{Y+Z} X = \nabla^*_Y X + \nabla^*_Z X$

(ii) $\nabla^*_{fY} X = f \nabla^*_Y X$

(iii) $\nabla^*_Z (X+Y) = \nabla^*_Z X + \nabla^*_Z Y$

(iv) $\nabla^*_Y (fX) = (Yf)X + f \nabla^*_Y X$

(II) リーマン多様体 (M, g) において[3]，計量 g に関する双対性:

$$Xg(Y,Z) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X Z)$$

を満たすアフィン接続のペア (∇, ∇^*) が与えられたとき，3 つの組 (g, ∇, ∇^*) を M の双対構造という[4]．M の双対構造に対して，次が成り立つ.

$$\nabla \text{ がリーマン接続（計量的）} \Longleftrightarrow \nabla = \nabla^* \text{ かつ捩率がゼロ}$$

つまり，$\nabla = \nabla^*$ であれば ∇ はリーマン接続（計量的）である．逆に，$\nabla \neq \nabla^*$ であれば，一般的には ∇ および ∇^* ともにリーマン接続（計量的）ではない．しかし，

$$\bar{\nabla} = \frac{1}{2}(\nabla + \nabla^*)$$

は常にリーマン接続（計量的）である.

証明：　双対アフィン接続の定義から，

$$Xg(Y,Z) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X Z)$$

[3]「多様体 M には内積 g が定義されている」の意味である.
[4]「多様体 M に内積 g と 2 つの接続 ∇ と ∇^* が定義されている」の意味である.

であり，また，Z と Y を入れ替えて，

$$Xg(Z,Y) = g(\nabla_X Z, Y) + g(Z, \nabla^*_X Y)$$

内積の順序を入れ替えて

$$Xg(Z,Y) = g(Z, \nabla^*_X Y) + g(\nabla_X Z, Y) = g(\nabla^*_X Y, Z) + g(Y, \nabla_X Z)$$

一番上の式と足して 2 で割れば,

$$Xg(Z,Y) = g(\bar{\nabla}_X Y, Z) + g(Y, \bar{\nabla}_X Z)$$

を得る．すなわち，$\bar{\nabla}$ は計量的である．

(III) $\nabla-$ 曲率 R がゼロであることと，∇^*- 曲率 R^* がゼロであることは同値である．

証明の概略：恒等式

$$[X,Y]g(Z,W) = XYg(Z,W) - YXg(Z,W)$$

において，まず左辺は $\nabla-$ 接続と ∇^*- 接続を用いて

$$[X,Y]g(Z,W) = g(\nabla_{[X,Y]}Z, W) + g(Z, \nabla^*_{[X,Y]}W)$$

となる．一方右辺の第 1 項は

$$XYg(Z,W) = X\left[Yg(Z,W)\right] = g(\nabla_X\nabla_Y Z, W) + g(\nabla_Y Z, \nabla^*_X W) + g(\nabla_X Z, \nabla^*_Y W) + g(Z, \nabla^*_X\nabla^*_Y W)$$

であり，右辺の第 2 項は

$$YXg(Z,W) = Y\left[Xg(Z,W)\right] = g(\nabla_Y\nabla_X Z, W) + g(\nabla_X Z, \nabla^*_Y W) + g(\nabla_Y Z, \nabla^*_X W) + g(Z, \nabla^*_Y\nabla^*_X W)$$

したがって，若干の計算の後,

$$-\left[g(\nabla_X\nabla_Y Z, W) - g(Z, \nabla_Y\nabla_X W) - g(\nabla_{[X,Y]}Z, W)\right] = g(Z, \nabla^*_X\nabla^*_Y W) - g(Z, \nabla^*_Y\nabla^*_X W) - g(Z, \nabla^*_{[X,Y]}W)$$

を得るので,

$$-g(R(X,Y,Z)Z, W) = g(Z, R^*(X,Y,Z)W)$$

が成り立つ．上式から，$R=0$ と $R^*=0$ は同値である．つまり，$R=0$ なら $R^*=0$ であり，$R^*=0$ なら $R=0$ である．

(IV) **統計多様体** 接続 ∇ が計量的である条件：

$$Xg(Y,Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$$

を思い出し，ここで,

$$(\nabla g)(X,Y,Z) = Xg(Y,Z) - g(\nabla_X Y, Z) - g(Y, \nabla_X Z) \tag{4.9}$$

を定義する．これが，接続 ∇ が「どのくらい計量的でないか」を測るテンソル量である．($(\nabla g)(X,Y,Z)$ は $(\nabla_X g)(Y,Z)$ と書く流儀もある．)

互いに双対な接続に対して，次の 4 条件のうち，任意の 2 条件を仮定すると，残りの 2 条件が成り立つ．

(1) ∇ の捩率はゼロである．

(2) ∇^* の捩率はゼロである．

(3) $(\nabla g)(X,Y,Z)$ は X,Y,Z に関して対称である．（コダッチの方程式）

(4) $\bar{\nabla}=\frac{1}{2}(\nabla+\nabla^*)$ は Riemann 接続である．

証明は Appendix の 4.7.2 節に記載.

多様体 M にアフィン接続 ∇ とリーマン計量 g が与えられていて，∇ とその双対接続 ∇^* の捩率がともにゼロであるとき (M,∇,g) を統計多様体と言う．すなわち，前述の 4 条件 (1)–(4) が満たされている（実際には任意の 2 条件が満たされていればよい）多様体を統計多様体と呼ぶ.（これは，狭い意味の統計多様体である．文献によっては「統計多様体」を広い意味で使い，確率分布族の空間を多様体と見たものをこう呼ぶ場合もある.）

4.3 アルファ接続

4.3.1 フィシャー計量

統計モデル（多様体）M に内積 g を導入する．すなわち，

$$\begin{aligned} g\,:\,\mathfrak{X}(M)\times\mathfrak{X}(M) &\quad\to\quad \mathbb{R} \\ g(X,Y)=\sum_{x\in\Omega}p(x|\boldsymbol{\theta})[X\log p(x|\boldsymbol{\theta})][Y\log p(x|\boldsymbol{\theta})] &= E_{\boldsymbol{\theta}}[(Xl_{\boldsymbol{\theta}})(Yl_{\boldsymbol{\theta}})] \qquad X,Y\,\in\mathfrak{X}(M)\end{aligned}$$

である．この内積から計算される計量はフィシャー計量と呼ばれ

$$g_{ij}=g(\partial_i,\partial_j)=\sum_{x\in\Omega}[\partial_i\log p(x|\boldsymbol{\theta})][\partial_j\log p(x|\boldsymbol{\theta})]p(x|\boldsymbol{\theta})=E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})] \tag{4.10}$$

となる．

統計多様体（すなわち確率分布族の空間）において不変なリーマン計量はフィシャー計量に限られる．

確率分布族の空間における不変性について：$\boldsymbol{\theta}$ をパラメータとする確率分布族 $p(x|\boldsymbol{\theta})$ を考える．今，確率変数 x を

$$y=f(x),\qquad x=f^{-1}(y)$$

と変換してみる．すると確率密度関数は

$$\widetilde{p}(y|\boldsymbol{\theta})=p(x|\boldsymbol{\theta})\frac{\partial f^{-1}(y)}{\partial y}=p(f^{-1}(y)|\boldsymbol{\theta})\frac{\partial f^{-1}(y)}{\partial y}$$

と変化する．しかし，多様体 $M=\{p(x|\boldsymbol{\theta})\}$ と $\widetilde{M}=\{\widetilde{p}(y|\boldsymbol{\theta})\}$ は，各点は同じ座標 $\boldsymbol{\theta}$ で指定され，分布は同じで確率変数の表現を変えただけである．したがって，M と $\widetilde{M}$ の幾何学構造は同じでなければならない．この要求を満たす計量はフィシャー計量に限られる．

4.3.2 アルファ接続の導入

X, Y, Z を多様体 M のベクトル場（$X, Y, Z \in \mathfrak{X}(M)$）として，$T(X, Y, X)$ を $(0, 3)$ 型のテンソル場とする．すなわち，$T(X, Y, X)$ は

$$T(X, Y, Z)\,;\, \mathfrak{X}(M) \times \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathbb{R}$$

である線形多重写像である．

$T(X, Y, Z)$ を用いると，任意のアフィン接続 ∇ から新しいアフィン接続 $\widetilde{\nabla}$ を

$$g(\widetilde{\nabla}_Y X, Z) = g(\nabla_Y X, Z) + T(X, Y, Z) \tag{4.11}$$

として作り出すことができる．上式の ∇ がアフィン接続であることはアフィン接続の満たすべき条件

(i) $\nabla_{\boldsymbol{X}+\boldsymbol{Y}}\boldsymbol{Z} = \nabla_{\boldsymbol{X}}\boldsymbol{Z} + \nabla_{\boldsymbol{Y}}\boldsymbol{Z}$

(ii) $\nabla_{f\boldsymbol{X}}\boldsymbol{Z} = f\nabla_{\boldsymbol{X}}\boldsymbol{Z}$

(iii) $\nabla_{\boldsymbol{X}}(\boldsymbol{Y}+\boldsymbol{Z}) = \nabla_{\boldsymbol{X}}\boldsymbol{Y} + \nabla_{\boldsymbol{X}}\boldsymbol{Z}$

(iv) $\nabla_{\boldsymbol{X}}(f\boldsymbol{Y}) = (\boldsymbol{X}f)\nabla_{\boldsymbol{X}}\boldsymbol{Y} + f\nabla_{\boldsymbol{X}}\boldsymbol{Y}$

を ∇ が満たすことを確かめればよい．$T(X, Y, X)$ が線形多重写像であることから，この事はほとんど自明である．

ここで，リーマン接続 ∇ に対して，新しい接続アフィン $\widetilde{\nabla}$ を

$$g(\widetilde{\nabla}_Y X, Z) = g(\nabla_Y X, Z) - \frac{\alpha}{2} T(X, Y, Z) \tag{4.12}$$

として定義する．ここで，$T(X, Y, Z)$ は式 (4.17) で定義した $(0, 3)$ 型テンソルである．$\frac{\alpha}{2}$ は定数であり，$\alpha = 0$ の場合 $\widetilde{\nabla} = \nabla$ である．すなわち，$\widetilde{\nabla}$ はリーマン接続に一致する．式 (4.12) で定義される新しいアフィン接続はアルファ接続と呼ばれる．この事を反映して新しい接続 $\widetilde{\nabla}$ を $\nabla^{(\alpha)}$ と表すことにする．すなわち，

$$g(\nabla^{(\alpha)}_Y X, Z) = g(\nabla_Y X, Z) - \frac{\alpha}{2} T(X, Y, Z) \tag{4.13}$$

と書く．

4.3.3 アルファ接続の接続係数の導出

この節ではアルファ接続の接続係数を導出する．式 (4.13) を再掲すると

$$g(\nabla^{(\alpha)}_Y X, Z) = g(\nabla_Y X, Z) - \frac{\alpha}{2} T(X, Y, Z)$$

に対して，$Y = \partial_i$, $X = \partial_j$, $Z = \partial_k$ を代入すれば

$$g(\nabla_Y X, Z) = g(\nabla_{\partial_i}\partial_j, \partial_k) = g({\Gamma^{\ell}}_{ij}\,\partial_\ell, \partial_k) = {\Gamma^{\ell}}_{ij}\, g(\partial_\ell, \partial_k) = {\Gamma^{\ell}}_{ij}\, g_{\ell k} = \Gamma_{ij,k} \tag{4.14}$$

$$g(\nabla^{(\alpha)}_Y X, Z) = g(\nabla^{(\alpha)}_{\partial_i}\partial_j, \partial_k) = \Gamma^{(\alpha)}_{ij,k} \tag{4.15}$$

である．ここで，$\Gamma_{ij,k}$ がリーマン接続 ∇ から計算される接続係数（第 1 種クリストッフェル記号），$\Gamma^{(\alpha)}_{ij,k}$ がアルファ接続 $\nabla^{(\alpha)}$ から計算される接続係数（第 1 種クリストッフェル記号）である．以下では，式 (4.13)，(4.14)，(4.15) から $\Gamma_{ij,k}$ と $\Gamma^{(\alpha)}_{ij,k}$ の関係を導出する．

まず，計量にフィッシャー計量

$$g_{ij} = \sum_{x\in\Omega}[\partial_i \log p(x|\boldsymbol{\theta})][\partial_j \log p(x|\boldsymbol{\theta})]p(x|\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})]$$

を仮定する．すると，まず，$\Gamma_{ij,k}$ は良く知られた公式：

$$\Gamma_{ij,k} = \frac{1}{2}\left(-g_{ij,k} + g_{jk,i} + g_{ki,j}\right) = \frac{1}{2}\left(-\partial_k g_{ij} + \partial_i g_{jk} + \partial_j g_{ki}\right)$$

がある．上式右辺を計算する．g_{ij} に ∂_k を作用させて，

$$\begin{aligned}\partial_k g_{ij} = \sum_{x\in\Omega}[\partial_k\partial_i \log p(x|\boldsymbol{\theta})][\partial_j \log p(x|\boldsymbol{\theta})]p(x|\boldsymbol{\theta}) + \sum_{x\in\Omega}[\partial_i \log p(x|\boldsymbol{\theta})][\partial_k\partial_j \log p(x|\boldsymbol{\theta})]p(x|\boldsymbol{\theta})\\ + \sum_{x\in\Omega}[\partial_i \log p(x|\boldsymbol{\theta})][\partial_j \log p(x|\boldsymbol{\theta})](\partial_k p(x|\boldsymbol{\theta}))\end{aligned} \tag{4.16}$$

を得る．ここで（よく使う）式変形

$$\partial_k p(x|\boldsymbol{\theta}) = [\partial_k \log p(x|\boldsymbol{\theta})]p(x|\boldsymbol{\theta})$$

を用いると，

$$\begin{aligned}\partial_k g_{ij} &= \sum_{x\in\Omega}[\partial_k\partial_i \log p(x|\boldsymbol{\theta})][\partial_j \log p(x|\boldsymbol{\theta})]p(x|\boldsymbol{\theta}) + \sum_{x\in\Omega}[\partial_i \log p(x|\boldsymbol{\theta})][\partial_k\partial_j \log p(x|\boldsymbol{\theta})]p(x|\boldsymbol{\theta})\\ &\quad + \sum_{x\in\Omega}[\partial_i \log p(x|\boldsymbol{\theta})][\partial_j \log p(x|\boldsymbol{\theta})][\partial_k \log p(x|\boldsymbol{\theta})]p(x|\boldsymbol{\theta})\\ &= E_{\boldsymbol{\theta}}[(\partial_k\partial_i \log p(x|\boldsymbol{\theta}))(\partial_j \log p(x|\boldsymbol{\theta}))] + E_{\boldsymbol{\theta}}[(\partial_i \log p(x|\boldsymbol{\theta}))(\partial_k\partial_j \log p(x|\boldsymbol{\theta})]\\ &\quad + E_{\boldsymbol{\theta}}[(\partial_i \log p(x|\boldsymbol{\theta}))(\partial_j \log p(x|\boldsymbol{\theta}))(\partial_k \log p(x|\boldsymbol{\theta}))]\\ &= E_{\boldsymbol{\theta}}[(\partial_k\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_k\partial_j l_{\boldsymbol{\theta}}] + E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})]\end{aligned}$$

となる．ここで，$l_{\boldsymbol{\theta}} = \log p(x|\boldsymbol{\theta})$ と書いた．

すると，

$$\begin{aligned}\partial_k g_{ij} &= E_{\boldsymbol{\theta}}[(\partial_k\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_k\partial_j l_{\boldsymbol{\theta}}] + E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})]\\ \partial_i g_{jk} &= E_{\boldsymbol{\theta}}[(\partial_i\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_j l_{\boldsymbol{\theta}})(\partial_i\partial_k l_{\boldsymbol{\theta}}] + E_{\boldsymbol{\theta}}[(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})(\partial_i l_{\boldsymbol{\theta}})]\\ \partial_j g_{ki} &= E_{\boldsymbol{\theta}}[(\partial_j\partial_k l_{\boldsymbol{\theta}})(\partial_i l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_k l_{\boldsymbol{\theta}})(\partial_j\partial_i l_{\boldsymbol{\theta}}] + E_{\boldsymbol{\theta}}[(\partial_k l_{\boldsymbol{\theta}})(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})]\end{aligned}$$

であるので，

$$\begin{aligned}2\Gamma_{ij,k} &= -\partial_k g_{ij} + \partial_i g_{jk} + \partial_j g_{ki}\\ &= -\left(E_{\boldsymbol{\theta}}[(\partial_k\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_k\partial_j l_{\boldsymbol{\theta}}] + E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})]\right)\\ &\quad + \left(E_{\boldsymbol{\theta}}[(\partial_i\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_j l_{\boldsymbol{\theta}})(\partial_i\partial_k l_{\boldsymbol{\theta}}] + E_{\boldsymbol{\theta}}[(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})(\partial_i l_{\boldsymbol{\theta}})]\right)\\ &\quad + \left(E_{\boldsymbol{\theta}}[(\partial_j\partial_k l_{\boldsymbol{\theta}})(\partial_i l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_k l_{\boldsymbol{\theta}})(\partial_j\partial_i l_{\boldsymbol{\theta}}] + E_{\boldsymbol{\theta}}[(\partial_k l_{\boldsymbol{\theta}})(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})]\right)\\ &= 2E_{\boldsymbol{\theta}}[(\partial_i\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}}) + (\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})]\end{aligned}$$

を得る．したがって，

$$\Gamma_{ij,k} = E_{\boldsymbol{\theta}}\left[(\partial_i\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}}) + \frac{1}{2}(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})\right] = E_{\boldsymbol{\theta}}\left[\left(\partial_i\partial_j l_{\boldsymbol{\theta}} + \frac{1}{2}(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})\right)(\partial_k l_{\boldsymbol{\theta}})\right]$$

である．

α 接続に関する接続係数を求めるのであるが，ここで $(0,3)$ 型のテンソル場として，

$$T(X,Y,Z)=\sum_{x\in\Omega}p(x|\boldsymbol{\theta})(X\log p(x|\boldsymbol{\theta}))(Y\log p(x|\boldsymbol{\theta}))(Z\log p(x|\boldsymbol{\theta}))$$
$$=E_{\boldsymbol{\theta}}[(X\log p(x|\boldsymbol{\theta}))(Y\log p(x|\boldsymbol{\theta}))(Z\log p(x|\boldsymbol{\theta}))] \quad (4.17)$$

を導入する．(統計多様体の不変性の要求を満たす $(0,3)$ 型のテンソルは，またしても上記のみであるのが理由である.）このとき，成分 T_{ijk} は

$$T_{ijk}=T(\partial_i,\partial_j,\partial_k)=E_{\boldsymbol{\theta}}\left[(\frac{\partial}{\partial\theta^i}\log p(x|\boldsymbol{\theta}))(\frac{\partial}{\partial\theta^j}\log p(x|\boldsymbol{\theta}))(\frac{\partial}{\partial\theta^k}\log p(x|\boldsymbol{\theta}))\right]$$

として求まる．

すると，結局 α 接続に関する接続係数.

$$\Gamma^{(\alpha)}_{ij,k}=\Gamma_{ij,k}-\frac{\alpha}{2}E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})]=E_{\boldsymbol{\theta}}\left[\left(\partial_i\partial_j l_{\boldsymbol{\theta}}+\frac{1-\alpha}{2}(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})\right)(\partial_k l_{\boldsymbol{\theta}})\right] \quad (4.18)$$

を得る．上式からすぐに，

$$\Gamma^{(\alpha)}_{ij,k}=\Gamma^{(\alpha)}_{ji,k}$$

が成り立つことがわかる．したがって，アルファ接続 $\nabla^{(\alpha)}$ は捩率がゼロである．

4.3.4 $\nabla^{(\alpha)}$ と $\nabla^{(-\alpha)}$ の双対性

統計多様体 M に対してフィシャー計量

$$g_{ij}=E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})]$$

を考え，接続 $\nabla^{(-\alpha)}$ と $\nabla^{(\alpha)}$ の関係を見る．式 (4.16) において既に示したように，

$$\partial_k g_{ij}=E_{\boldsymbol{\theta}}[(\partial_k\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})]+E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_k\partial_j l_{\boldsymbol{\theta}})]++E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})] \quad (4.19)$$

が成り立つ．ところで，

$$g(\nabla^{(\alpha)}_{\partial_k}\partial_i,\partial_j)=\Gamma^{(\alpha)}_{ki,j} \qquad g(\partial_i,\nabla^{(-\alpha)}_{\partial_k}\partial_j)=\Gamma^{(-\alpha)}_{kj,i}$$

であり，アルファ接続のクリストッフェル記号は式 (4.18):

$$\Gamma^{(\alpha)}_{ij,k}=E_{\boldsymbol{\theta}}\left[\left(\partial_i\partial_j l_{\boldsymbol{\theta}}+\frac{1-\alpha}{2}(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})\right)(\partial_k l_{\boldsymbol{\theta}})\right]$$

で与えられるので，結局，

$$\begin{aligned}g(\nabla^{(\alpha)}_{\partial_k}\partial_i,\partial_j)+g(\partial_i,\nabla^{(-\alpha)}_{\partial_k}\partial_j)&=\Gamma^{(\alpha)}_{ki,j}+\Gamma^{(-\alpha)}_{kj,i}\\
&=E_{\boldsymbol{\theta}}\left[\left(\partial_k\partial_i l_{\boldsymbol{\theta}}+\frac{1-\alpha}{2}(\partial_k l_{\boldsymbol{\theta}})(\partial_i l_{\boldsymbol{\theta}})\right)(\partial_j l_{\boldsymbol{\theta}})\right]+E_{\boldsymbol{\theta}}\left[\left(\partial_k\partial_j l_{\boldsymbol{\theta}}+\frac{1+\alpha}{2}(\partial_k l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})\right)(\partial_i l_{\boldsymbol{\theta}})\right]\\
&=E_{\boldsymbol{\theta}}[(\partial_k\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})]+E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_k\partial_j l_{\boldsymbol{\theta}})]++E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})]=\partial_k g_{ij}\end{aligned}$$

が成り立つ．すなわち，$\nabla^{(-\alpha)}$ は $\nabla^{(\alpha)}$ の双対接続である．したがって，$(M,\nabla^{(\alpha)},g)$ は統計多様体である．つまり，統計モデル M はフィシャー計量とアルファ接続を付与することにより統計多様体になる．

4.4　双対平坦性とルジャンドル変換

4.4.1　双対アフィン座標

ここからは，多様体の双対平坦性について議論する．

定義：　双対構造 (g, ∇, ∇^*) を持つ多様体 M において，∇ に関する曲率 R も捩率 T もともにゼロとなり，かつ ∇^* に関する曲率 R^* と捩率 T^* がゼロとなるとき，M **は双対平坦**であるという．(若干冗長である．要は，双対構造 (g, ∇, ∇^*) を持つ統計多様体 M において，$R = 0$ であれば M は双対平坦である．)

平坦な多様体上では局所アフィン座標系（接続が恒等的にゼロとなる座標系）が存在する．接続 ∇ と ∇^* について同時に平坦な多様体を考えているので，局所アフィン座標系もそれぞれの接続ごとに 2 種類存在する．そこで，$(U; x^1, \ldots, x^n)$ を ∇ に関する局所アフィン座標系，$(V; y^1, \ldots, y^n)$ を ∇^* に関する局所アフィン座標系とする．

定理 1　双対平坦な多様体 M では，各点の周りで

$$g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^j}\right) = \delta_{ij} \tag{4.20}$$

を満たす局所 $\nabla-$ アフィン座標系 (x^i) と，局所 ∇^*- アフィン座標系 (y^j) を取ることができる．この特性を持った (x^i) と (y^j) を**双対アフィン座標系**と呼ぶ.

証明は 4.7.3 節に記載.

記号法の変更説明：これ以降，双対座標を $(x^1, \ldots, x^n)$，$(y_1, \ldots, y_n)$ と書く．片方の座標は下付き添え字を用いて書く.）それに伴い，

$$\frac{\partial}{\partial x^i} = \partial_i \qquad \frac{\partial}{\partial y_i} = \partial^i$$

とする．式 (4.20) の関係は

$$g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y_j}\right) = g(\partial_i, \partial^j) = \delta_i^j$$

と表される．

4.4.2　双対アフィン座標とルジャンドル変換

まず，以下の補題を証明する．

補題：双対アフィン座標系 $\{\theta^i, \eta_i\}$ に対する計量は

$$g_{ij} = g(\partial_i, \partial_j) = g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\theta^j}\right) \qquad g^{ij} = g(\partial^i, \partial^j) = g\left(\frac{\partial}{\partial\eta_i}, \frac{\partial}{\partial\eta_j}\right)$$

に等しい．計量に関して

$$g_{ij} = \frac{\partial\eta_j}{\partial\theta^i} = \frac{\partial\eta_i}{\partial\theta^j}$$
$$g^{ij} = \frac{\partial\theta^j}{\partial\eta_i} = \frac{\partial\theta^i}{\partial\eta_j}$$
$$g_{ij}g^{jk} = \delta_i^k$$

が成り立つ．

証明：

$$\partial_i = \frac{\partial}{\partial\theta^i} = \frac{\partial\eta_k}{\partial\theta^i}\frac{\partial}{\partial\eta_k} = \frac{\partial\eta_k}{\partial\theta^i}\partial^k$$

を用いれば

$$g_{ij} = g(\partial_i, \partial_j) = g(\frac{\partial\eta_k}{\partial\theta^i}\partial^k, \partial_j) = \frac{\partial\eta_k}{\partial\theta^i}g(\partial^k, \partial_j) = \frac{\partial\eta_k}{\partial\theta^i}\delta_j^k = \frac{\partial\eta_j}{\partial\theta^i} \tag{4.21}$$

が成り立つ．ここで，g_{ij} は座標変換 $(\theta^i) \to (\eta_j)$ の座標変換のヤコビ行列 $\left(\frac{\partial\eta_j}{\partial\theta^i}\right)$ である．また，計量は対称であるので（内積は順序によらないので）

$$g_{ij} = g_{ji} = \frac{\partial\eta_i}{\partial\theta^j}$$

も成り立つ．

また，

$$\partial^i = \frac{\partial}{\partial\eta_i} = \frac{\partial\theta^k}{\partial\eta_i}\frac{\partial}{\partial\theta^k} = \frac{\partial\theta^k}{\partial\eta_i}\partial_k$$

したがって，

$$g^{ij} = g(\partial^i, \partial^j) = \frac{\partial\theta^k}{\partial\eta_i}g(\partial_k, \partial^j) = \frac{\partial\theta^k}{\partial\eta_i}\delta_k^j = \frac{\partial\theta^j}{\partial\eta_i} \tag{4.22}$$

である．g^{ij} は座標変換 $(\eta_i) \to (\theta^j)$ のヤコビ行列 $\left(\frac{\partial\theta^j}{\partial\eta_i}\right)$ である．また，計量の対称性から

$$g^{ij} = g^{ji} = \frac{\partial\theta^i}{\partial\eta_j}$$

も成り立つ．最後に

$$g_{ij}g^{jk} = \frac{\partial\eta_j}{\partial\theta^i}\frac{\partial\theta^k}{\partial\eta_j} = \delta_i^k$$

は，ヤコビ行列とその逆行列を乗じているので単位行列になる．

定理 I (M, g, ∇, ∇^*) を n 次元双対平坦な多様体，$\theta = (\theta^1, \ldots, \theta^n)$，$\eta = (\eta_1, \ldots, \eta_n)$ を M の双対アフィン座標系とする．g_{ij} と g^{ij} は

$$g_{ij} = g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\theta^j}\right) \qquad g^{ij} = g\left(\frac{\partial}{\partial\eta_i}, \frac{\partial}{\partial\eta_j}\right)$$

として計算できる．すると，ある凸関数 $\psi(\theta)$ と $\varphi(\eta)$ が存在し，

$$g_{ij} = \frac{\partial^2\psi}{\partial\theta^i\partial\theta^j} \qquad g_{ij} = \frac{\partial^2\varphi}{\partial\eta_i\partial\eta_j}$$

と表され，また，

$$\theta^i = \frac{\partial\varphi}{\partial\eta_i} \qquad \eta_i = \frac{\partial\psi}{\partial\theta^i} \qquad \psi(\theta) + \varphi(\eta) = \theta^k\eta_k$$

が成り立つ．ここで，$\psi(\theta)$，$\varphi(\eta)$ はポテンシャルと呼ばれる[5]．つまり，θ^i はポテンシャル $\varphi(\eta)$ の η_i に関する勾配として，η^j はポテンシャル $\psi(\theta)$ の θ_j に関する勾配として求まるのである．

[5] θ と η はそれぞれ $\theta^1, \ldots, \theta^n$ と $\eta_1, \ldots, \eta_n$ を集合的に表したもので，$\psi(\theta) = \psi(\theta^1, \ldots, \theta^n)$，$\varphi(\eta) = \varphi(\eta_1, \ldots, \eta_n)$ の意味である．

証明：　関係式：

$$\frac{\partial \eta_i}{\partial \theta^j} = \frac{\partial \eta_j}{\partial \theta^i}$$

は連立微分方程式

$$\eta_i = \frac{\partial \psi}{\partial \theta^i} \qquad i = 1, \ldots, n$$

において，ポテンシャル関数 ψ（凸関数 ψ）が解として存在することの条件である．したがって，式 (4.21) を用いて，

$$g_{ij} = \frac{\partial \eta_i}{\partial \theta^j} = \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}$$

を得る．また，関係式：

$$\frac{\partial \theta^j}{\partial \eta_i} = \frac{\partial \theta^i}{\partial \eta_j}$$

は連立微分方程式

$$\theta^i = \frac{\partial \varphi}{\partial \eta_i} \qquad i = 1, \ldots, n$$

において，ポテンシャル関数 φ（凸関数 φ）が解として存在することの条件である．したがって，式 (4.22) を用いて

$$g^{ij} = \frac{\partial \theta^i}{\partial \eta_j} = \frac{\partial^2 \varphi}{\partial \eta_i \partial \eta_j}$$

を得る．

ここで，$\psi + \varphi - \theta^k \eta_k$ の全微分を計算すると

$$d\psi + d\varphi - (d\theta^k)\eta_k - \theta^k(d\eta_k) = \frac{\partial \psi}{\partial \theta^i} d\theta^i + \frac{\partial \varphi}{\partial \eta_i} d\eta_i - \eta_i d\theta^i - \theta^i d\eta_i = \eta_i d\theta^i + \theta^i d\eta_i - \eta_i d\theta^i - \theta^i d\eta_i = 0$$

したがって，$\psi + \varphi - \theta^k \eta_k$ は定数関数である．ポテンシャル関数には定数項の任意性があるので，ポテンシャル関数に定数を加えることにより

$$\psi + \varphi - \theta^k \eta_k = 0 \tag{4.23}$$

とすることができる．

定理 II ポテンシャル関数は $p, q \in M$ に対して

$$\varphi(\eta(p)) = \max_{q \in M} \left(\theta^i(q) \eta_i(p) - \psi(\theta(q)) \right) \tag{4.24}$$

$$\psi(\theta(p)) = \max_{q \in M} \left(\eta_i(q) \theta^i(p) - \varphi(\eta(q)) \right) \tag{4.25}$$

と表される．

証明 [式 (4.24)]**：**p は固定し，q で微分すれば，

$$\begin{aligned} \frac{d}{dq} \left(\theta^i(q) \eta_i(p) - \psi(\theta(q)) \right) &= \left(\frac{d\theta^i(q)}{dq} \eta_i(p) - \frac{d\psi(\theta(q))}{dq} \right) \\ &= \left(\eta_i(p) - \frac{\partial \psi(\theta(q))}{\partial \theta^i(q)} \right) \frac{d\theta^i(q)}{dq} = \left(\eta_i(p) - \eta_i(q) \right) \frac{d\theta^i(q)}{dq} \end{aligned}$$

である．したがって，式 (4.24) の max はすべての i で $\eta_i(p) = \eta_i(q)$，すなわち，$p = q$ のときに達成される．その最大値は

$$\varphi(\eta(p)) = \theta^i(p) \eta_i(p) - \psi(\theta(p))$$

を満たす．これは式 (4.23) が成り立つ条件である．式 (4.25) も同様に証明できる．

4.5　指数型分布族と双対平坦性

双対平坦性を，確率分布を指数型分布族に限定して議論する．

4.5.1　指数型分布族

指数型分布族とは，確率（密度）分布が

$$p(x|\boldsymbol{\theta}) = \exp\left(\theta^i u_i(x) - \psi(\boldsymbol{\theta}) + c(x)\right) = \exp\left(\boldsymbol{\theta}\cdot\boldsymbol{u}(x) - \psi(\boldsymbol{\theta}) + c(x)\right) \tag{4.26}$$

の形を持つ確率分布の族を言う．ここで，確率分布のパラメータは $\boldsymbol{\theta} = (\theta^1, \ldots, \theta^n)$ であり，x が確率変数である．x は多変数の場合にはベクトル $\boldsymbol{x}$ とする．$u_i(x)$ は確率変数の関数であり，$\boldsymbol{u}(x) = (u_1(x), \ldots, u_n(x))$ である統計多様体を，

$$M = \{p_{\boldsymbol{\theta}} \,|\, p_{\boldsymbol{\theta}} = \exp\left(\theta^i u_i(x) - \psi(\boldsymbol{\theta}) + c(x)\right), \boldsymbol{\theta} \in \Theta\}$$

と表す．確率分布は積分して 1 であるので，

$$1 = \int p(x|\boldsymbol{\theta})dx = \int \exp\left(\boldsymbol{\theta}\cdot\boldsymbol{u} - \psi(\boldsymbol{\theta})\right)dx \quad \to \quad \int \exp\left(\boldsymbol{\theta}\cdot\boldsymbol{u}\right)dx = \exp(\psi(\boldsymbol{\theta}))$$

であるので，

$$\psi(\boldsymbol{\theta}) = \log \int \exp\left(\boldsymbol{\theta}\cdot\boldsymbol{u}(x)\right)dx \tag{4.27}$$

を得る．

この $\psi(\boldsymbol{\theta})$ は凸関数である．これは，以降の議論に重要なポイントである．

証明： 2 次の微分係数 $\dfrac{\partial^2}{\partial\theta^i\partial\theta^j}\psi(\boldsymbol{\theta})$ を (i,j) 要素に持つ行列を Hesse 行列と呼ぶ．Hesse 行列が正定値行列なら $\psi(\boldsymbol{\theta})$ は凸関数であるので，これを調べる．

$$\begin{aligned}
\frac{\partial^2}{\partial\theta^i\partial\theta^j}\psi(\boldsymbol{\theta}) &= \frac{\partial}{\partial\theta^i}\frac{\int u_j(x)\exp\left(\boldsymbol{\theta}\cdot\boldsymbol{u}(x)\right)dx}{\int \exp\left(\boldsymbol{\theta}\cdot\boldsymbol{u}(x)\right)dx} \\
&= \frac{\int u_i(x)u_j(x)\exp\left(\boldsymbol{\theta}\cdot\boldsymbol{u}(x)\right)dx \int \exp\left(\boldsymbol{\theta}\cdot\boldsymbol{u}(x)\right)dx - \int u_i(x)\exp\left(\boldsymbol{\theta}\cdot\boldsymbol{u}(x)\right)dx \int u_j(x)\exp\left(\boldsymbol{\theta}\cdot\boldsymbol{u}(x)\right)dx}{\left(\int \exp\left(\boldsymbol{\theta}\cdot\boldsymbol{u}(x)\right)dx\right)^2} \\
&= \frac{\int u_i(x)u_j(x)p(x|\boldsymbol{\theta})dx \int p(x|\boldsymbol{\theta})dx - \int u_i(x)p(x|\boldsymbol{\theta})dx \int u_j(x)p(x|\boldsymbol{\theta})dx}{\left(\int p(x|\boldsymbol{\theta})dx\right)^2} \\
&= E_{\boldsymbol{\theta}}[u_i(x)u_j(x)] - E_{\boldsymbol{\theta}}[u_i(x)]E_{\boldsymbol{\theta}}[u_j(x)] = \mathrm{Cov}(u_i(x), u_j(x))
\end{aligned}$$

となり，$\psi(\boldsymbol{\theta})$ の Hesse 行列は $u_i(x)$ と $u_j(x)$ の共分散行列に等しい．共分散行列は正定値行列であるので，$\psi(\boldsymbol{\theta})$ は凸関数である．

指数型分布族の簡単な例を挙げる．

スカラー（1 次元）正規分布

$$p(x|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

の場合，

$$\log p(x|\boldsymbol{\theta}) = \log\left[\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]\right] = \frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \left(\frac{\mu^2}{2\sigma^2} + \log(\sqrt{2\pi}\sigma)\right)$$

であるので，

$$c(x) = 0, \quad \theta^1 = \frac{\mu}{\sigma^2}, \quad u_1(x) = x, \quad \theta^2 = -\frac{1}{2\sigma^2}, \quad u_2(x) = x^2$$

であり，

$$\psi(\theta^1, \theta^2) = -\frac{(\theta^1)^2}{4\theta^2} + \frac{1}{2}\log\left(-\frac{\pi}{\theta^2}\right) = -\frac{(\theta^1)^2}{4\theta^2} - \frac{1}{2}\log\left(-\theta^2\right) \tag{4.28}$$

である.（定数項は無視した.）

4.5.2 指数型分布族における曲率

式 (4.18) から，指数型分布という確率分布を持つ族に対して，$\alpha = 1$ とした場合に，アルファ接続 $\nabla^{(1)}$ の曲率テンソルがゼロとなることを示す.

指数型分布族の確率分布は式 (4.26) より，

$$p(x|\boldsymbol{\theta}) = \exp\left(\theta^i u_i(x) - \psi(\boldsymbol{\theta}) + c(x)\right)$$

である．したがって，

$$\partial_j l_{\boldsymbol{\theta}} = \frac{\partial}{\partial\theta^j}\left(\theta^i u_i(x) - \psi(\boldsymbol{\theta}) + c(x)\right) = u_i(x) - \frac{\partial\psi(\boldsymbol{\theta})}{\partial\theta^j}$$

であり，さらに，

$$\partial_i\partial_j l_{\boldsymbol{\theta}} = \frac{\partial}{\partial\theta^i}\left(u_j(x) - \frac{\partial\psi(\boldsymbol{\theta})}{\partial\theta^j}\right) = -\frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}(\boldsymbol{\theta})$$

を得る.

したがって，$\alpha = 1$ の場合のアルファ接続係数 $\Gamma^{(1)}_{ij,k}$ は，

$$\Gamma^{(1)}_{ij,k} = E_{\boldsymbol{\theta}}\left[(\partial_i\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})\right] = -E_{\boldsymbol{\theta}}\left[\left(\frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}(\boldsymbol{\theta})\right)(\partial_k l_{\boldsymbol{\theta}})\right] = -\frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}(\boldsymbol{\theta})E_{\boldsymbol{\theta}}\left[(\partial_k l_{\boldsymbol{\theta}})\right]$$

であるが，

$$E_{\boldsymbol{\theta}}\left[(\partial_k l_{\boldsymbol{\theta}})\right] = \sum_{x\in\Omega}\left(\frac{\partial}{\partial\theta^k}\log p(x|\boldsymbol{\theta})\right)p(x|\boldsymbol{\theta}) = \sum_{x\in\Omega}\left(\frac{\partial}{\partial\theta^k}p(x|\boldsymbol{\theta})\right) = \frac{\partial}{\partial\theta^k}\sum_{x\in\Omega}p(x|\boldsymbol{\theta}) = \frac{\partial}{\partial\theta^k}1 = 0$$

であるので，$\Gamma^{(1)}_{ij,k} = 0$ となる．つまり，このとき $\boldsymbol{\theta}$ はアフィン座標系であり，アルファ接続 $\nabla^{(1)}$ は平坦である[6].

なお，$\alpha = 1$ であるアルファ接続は指数型接続，あるいは e-接続と呼ばれる.（ちなみに，本ノートでは議論しないが，$\alpha = -1$ であるアルファ接続は混合型接続，あるいは m-接続と呼ばれる.）

---

[6] [3] には，アルファ接続に対するリーマン曲率テンソル $R^{(\alpha)}_{ijkl}$ とリーマン接続に対する曲率テンソル R_{ijkl} の間に

$$R^{(\alpha)}_{ijkl} = \frac{1-\alpha^2}{2}R_{ijkl}$$

の関係があることが述べられている．上式を用いれば $\alpha = \pm 1$ で $R^{(\alpha)}_{ijkl} = 0$ となることはより直接的に示せる.

4.5.3 ここまでの議論のまとめ

ここまでの議論を一旦まとめてみる．

確率分布族の多様体 M に対して，フィシャー計量

$$g_{ij} = E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})]$$

と，式 (4.17) で表される $(0,3)$ 型テンソル

$$T(X,Y,Z) = E_{\boldsymbol{\theta}}[(X\log p(x|\boldsymbol{\theta}))(Y\log p(x|\boldsymbol{\theta}))(Z\log p(x|\boldsymbol{\theta}))] \tag{4.17}$$

を導入する．ここで，アルファ接続 $\nabla^{(\alpha)}$ を

$$g(\nabla^{(\alpha)}_Y X, Z) = g(\nabla_Y X, Z) - \frac{\alpha}{2}T(X,Y,Z)$$

として導入すれば，アルファ接続による接続係数 $\Gamma^{(\alpha)}_{ij,k}$ は

$$\Gamma^{(\alpha)}_{ij,k} = \Gamma_{ij,k} - \frac{\alpha}{2}E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})]$$

として求まる．ここで，$\Gamma_{ij,k}$ はフィシャー計量を用いたリーマン接続による接続係数である．この式からすぐにわかる通り，アルファ接続による接続係数は

$$\Gamma^{(\alpha)}_{ij,k} = \Gamma^{(\alpha)}_{ji,k}$$

を満たすので，振率がゼロである．また，

$$\Gamma^{(\alpha)}_{ij,k} + \Gamma^{(-\alpha)}_{ij,k} = \partial_k g_{ij}$$

を満たすので，$\nabla^{(\alpha)}$ と $\nabla^{(-\alpha)}$ は互いに双対接続の関係にある．したがって，多様体 $(M, \nabla^{(\alpha)}, g)$ は統計多様体となる．

さらに，指数分布族の多様体（統計モデル）

$$M = \{p_{\boldsymbol{\theta}} \,|\, p_{\boldsymbol{\theta}} = \exp\left(\theta^i u_i(x) - \psi(\boldsymbol{\theta}) + c(x)\right), \boldsymbol{\theta} \in \Theta\}$$

に話を限定すれば，$\alpha = 1$ の場合に，恒等的に

$$\Gamma^{(1)}_{ij,k} = 0$$

である．つまり，$\boldsymbol{\theta} = (\theta^1, \ldots, \theta^n)$ はアフィン座標系であり，アフィン座標系が存在するのでリーマン曲率テンソル $R^{(1)}$ がゼロとなることがわかる．また，双対接続 $\nabla^{(-1)}$ に対してもリーマン曲率テンソル $R^{(-1)}$ はゼロとなるので，結局，指数分布族の多様体 M は双対平坦である．

統計モデル M は，フィシャー計量 g_{ij} とアルファ接続を用いた双対構造 $\nabla^{(-\alpha)}$，$\nabla^{(\alpha)}$ のもとで統計多様体である．さらに，M を指数分布族の統計モデルに限定すれば，$\alpha = 1$ において M は双対平坦となる．

4.5.4 指数分布族に対する双対理論

指数分布族の多様体は $\alpha = 1$ において M は双対平坦となる．このとき $\boldsymbol{\theta} = (\theta^1, \ldots, \theta^n)$ がアフィン座標系である．式 (4.27)

$$\psi(\boldsymbol{\theta}) = \log \int \exp\left(\boldsymbol{\theta} \cdot \boldsymbol{u}(x)\right) dx$$

の $\psi(\boldsymbol{\theta})$ を用いて，$\psi(\boldsymbol{\theta})$ の 2 階偏微分を計算してみれば,

$$\frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j} = E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})] = g_{ij} \tag{4.29}$$

である.

証明：証明は簡単である.

$$\log p(x|\boldsymbol{\theta}) = \theta^i u_i(x) - \psi(\boldsymbol{\theta}) + c(x)$$

の両辺を θ^i と θ^j で 2 回微分すれば，右辺で残るのは $\psi(\boldsymbol{\theta})$ の項だけである．すなわち,

$$\frac{\partial^2}{\partial \theta^i \partial \theta^j} \log p(x|\boldsymbol{\theta}) = -\frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}$$

である．したがって,

$$\frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j} = E_{\boldsymbol{\theta}}[\frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}] = -E_{\boldsymbol{\theta}}[\frac{\partial^2}{\partial \theta^i \partial \theta^j} \log p(x|\boldsymbol{\theta})] = E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})] = g_{ij}$$

である.

追記： ここで，最後の等号：

$$-E_{\boldsymbol{\theta}}[\frac{\partial^2}{\partial \theta^i \partial \theta^j} \log p(x|\boldsymbol{\theta})] = E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})]$$

の成立について追加の説明を行う．まず,

$$E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})] = \sum_{x \in \Omega} \left(\frac{\partial}{\partial \theta^i} \log p(x|\boldsymbol{\theta}) \right) \left(\frac{\partial}{\partial \theta^j} \log p(x|\boldsymbol{\theta}) \right) p(x|\boldsymbol{\theta}) = \sum_{x \in \Omega} \left(\frac{\partial}{\partial \theta^i} \log p(x|\boldsymbol{\theta}) \right) \frac{\partial}{\partial \theta^j} p(x|\boldsymbol{\theta}) \tag{4.30}$$

である．ところで,

$$E_{\boldsymbol{\theta}} \left[\frac{\partial}{\partial \theta^i} \log p(x|\boldsymbol{\theta}) \right] = \sum_{x \in \Omega} \left(\frac{\partial}{\partial \theta^i} p(x|\boldsymbol{\theta}) \right) = \frac{\partial}{\partial \theta^i} \sum_{x \in \Omega} p(x|\boldsymbol{\theta}) = 0$$

であるので,

$$\frac{\partial}{\partial \theta^j} E \left[\frac{\partial}{\partial \theta^i} \log p(x|\boldsymbol{\theta}) \right] = \frac{\partial}{\partial \theta^j} \sum_{x \in \Omega} \left(\frac{\partial}{\partial \theta^i} \log p(x|\boldsymbol{\theta}) \right) p(x|\boldsymbol{\theta}) = 0$$

である．したがって,

$$\begin{aligned} &\frac{\partial}{\partial \theta^j} \sum_{x \in \Omega} \left(\frac{\partial}{\partial \theta^i} \log p(x|\boldsymbol{\theta}) \right) p(x|\boldsymbol{\theta}) \\ &\qquad = \sum_{x \in \Omega} \left(\frac{\partial^2}{\partial \theta^j \partial \theta^i} \log p(x|\boldsymbol{\theta}) \right) p(x|\boldsymbol{\theta}) + \sum_{x \in \Omega} \left(\frac{\partial}{\partial \theta^i} \log p(x|\boldsymbol{\theta}) \right) \frac{\partial}{\partial \theta^j} p(x|\boldsymbol{\theta}) = 0 \end{aligned}$$

であるので，結局,

$$\sum_{x \in \Omega} \left(\frac{\partial}{\partial \theta^i} \log p(x|\boldsymbol{\theta}) \right) \frac{\partial}{\partial \theta^j} p(x|\boldsymbol{\theta}) = -\sum_{x \in \Omega} \left(\frac{\partial^2}{\partial \theta^j \partial \theta^i} \log p(x|\boldsymbol{\theta}) \right) p(x|\boldsymbol{\theta}) = -E_{\boldsymbol{\theta}} \left[\frac{\partial^2}{\partial \theta^j \partial \theta^i} \log p(x|\boldsymbol{\theta}) \right]$$

が成り立つ．式 (4.30) の最右辺に代入して,

$$E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})] = -E_{\boldsymbol{\theta}} \left[\frac{\partial^2}{\partial \theta^j \partial \theta^i} \log p(x|\boldsymbol{\theta}) \right] = E_{\boldsymbol{\theta}} \left[\frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j} \right] = \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}$$

を得る．

すなわち，式 (4.29) の右辺はフィシャー計量 g_{ij} に等しく，したがって，定理 I における双対アフィン座標 $\boldsymbol{\theta}$ に対するポテンシャル $\psi(\boldsymbol{\theta})$ は指数関数族の右辺に含まれている $\psi(\boldsymbol{\theta})$ に等しい．(等しいのでそもそも同じ記号 ψ を用いている訳である.）

興味深いのは．計量がフィシャー行列でなければならないことが，指数分布族であれば自然に導かれることである．もっと一般的に，確率分布族の多様体で計量がフィシャー行列でなければならないことは確率分布が備えているべき不変性からも導かれるが，この証明は難解である．

もう片方の双対アフィン座標 η_i は

$$\eta_i = \frac{\partial \psi(\boldsymbol{\theta})}{\partial \theta^i} = \frac{\partial}{\partial \theta^i} \log \int \exp\left(\boldsymbol{\theta} \cdot \boldsymbol{u}(x)\right) dx = \frac{\dfrac{\partial}{\partial \theta^i} \displaystyle\int \exp\left(\boldsymbol{\theta} \cdot \boldsymbol{u}(x)\right) dx}{\int \exp\left(\boldsymbol{\theta} \cdot \boldsymbol{u}(x)\right) dx}$$
$$= \int u_i(x) \exp\left(\boldsymbol{\theta} \cdot \boldsymbol{u}(x) - \psi(\boldsymbol{\theta})\right) dx = \int u_i(x) p(x|\boldsymbol{\theta}) dx = E_{\boldsymbol{\theta}}[u_i(x)]$$

として求まる．η_i は $u_i(x)$ の期待値に等しい．η_i は，$\alpha = -1$ の場合のアルファ接続 $\nabla^{(-\alpha)}$ に対するアフィン座標系である．$\eta_i = E_{\boldsymbol{\theta}}[u_i(x)]$ を期待値座標系と呼ぶ.

$\boldsymbol{\eta}$ に対するポテンシャル $\varphi(\boldsymbol{\eta})$ を求める．まず，

$$\log p(x|\boldsymbol{\theta}) = \theta^i u_i(x) - \psi(\boldsymbol{\theta}) + c(x)$$

の両辺の期待値を取れば，

$$E_{\boldsymbol{\theta}}[\log p(x|\boldsymbol{\theta})] = \theta^i E(u_i(x)) - \psi(\boldsymbol{\theta}) + c(x) = \theta^i \eta_i - \psi(\boldsymbol{\theta}) + c(x)$$

であるので，

$$\theta^i = \frac{\partial}{\partial \eta_i} E_{\boldsymbol{\theta}}[\log p(x|\boldsymbol{\theta})]$$

を得る．したがって，

$$\varphi(\eta) = E_{\boldsymbol{\theta}}[\log p(x|\boldsymbol{\theta})]$$

を得る．つまり，$\varphi(\eta)$ は確率のエントロピーの符号を変えたものになっている．ポテンシャルは定数の加算に対する任意性があるので，

$$\varphi(\eta) = E_{\boldsymbol{\theta}}[\log p(x|\boldsymbol{\theta})] - E_{\boldsymbol{\theta}}[c(x)] \tag{4.31}$$

とすることももちろん可能である．

4.5.5　スカラー（1 次元）正規分布での双対理論

双対座標とポテンシャル

スカラー（1 次元）正規分布

$$p(x|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

の場合，再掲すると

$$\log p(x|\boldsymbol{\theta}) = \log\left[\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]\right] = \frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \left(\frac{\mu^2}{2\sigma^2} + \log(\sqrt{2\pi}\sigma)\right)$$

であるので，

$$c(x) = 0, \quad \theta^1 = \frac{\mu}{\sigma^2}, \quad u_1(x) = x, \quad \theta^2 = -\frac{1}{2\sigma^2}, \quad u_2(x) = x^2$$

であり，

$$\psi(\theta^1, \theta^2) = -\frac{(\theta^1)^2}{4\theta^2} + \frac{1}{2}\log\left(-\frac{\pi}{\theta^2}\right) = -\frac{(\theta^1)^2}{4\theta^2} - \frac{1}{2}\log\left(-\theta^2\right) \tag{4.32}$$

である.（定数項は無視した.）上に示す $\boldsymbol{\theta} = (\theta^1, \theta^2)$ の取り方は，指数関数族の定義に使った θ^i の取り方と同じであり，自然パラメータあるいは正準パラメータと呼ばれる.

ここで，$\boldsymbol{\eta}$ 座標を求めると，

$$\eta_1 = \frac{\partial\psi}{\partial\theta^1} = \mu$$

$$\eta_2 = \frac{\partial\psi}{\partial\theta^2} = \frac{1}{4}\left(\frac{\theta^1}{\theta^2}\right)^2 + \frac{1}{2\theta^2} = \mu^2 + \sigma^2$$

である．また，ポテンシャル $\varphi(\boldsymbol{\eta})$ は

$$\varphi(\boldsymbol{\eta}) = E_{\boldsymbol{\theta}}[\log p(x|\boldsymbol{\theta})] = E_{\boldsymbol{\theta}}[-\log(\sqrt{2\pi}\sigma) - \frac{(x-\mu)^2}{2\sigma^2}] = -\log\sigma - \frac{E_{\boldsymbol{\theta}}[(x-\mu)^2]}{2\sigma^2} = -\log\sigma - \frac{\sigma^2}{2\sigma^2} = -\log\sigma$$

ところで，

$$\sigma^2 = \eta_2 - \eta_1^2$$

であるので，

$$\varphi(\eta_1, \eta_2) = -\frac{1}{2}\log(\eta_2 - \eta_1^2)$$

を得る．ここで，ちなみに，

$$\frac{\partial\varphi}{\partial\eta_1} = -\frac{1}{2}\frac{-2\eta_1}{\eta_2 - \eta_1^2} = \frac{\mu}{\sigma^2} = \theta^1$$

$$\frac{\partial\varphi}{\partial\eta_2} = -\frac{1}{2}\frac{1}{\eta_2 - \eta_1^2} = -\frac{1}{\sigma^2} = \theta^2$$

を得る．

$$\sigma^2 = -\frac{1}{2\theta^2} \qquad \mu = -\frac{\theta^1}{2\theta^2}$$

となることを参考にして，計量を計算してみると，

$$g_{11} = \frac{\partial^2\psi}{\partial(\theta^1)^2} = -\frac{1}{2\theta^2} = \sigma^2$$

$$g_{12} = g_{21} = \frac{\partial^2\psi}{\partial\theta^2\partial\theta^1} = -\frac{1}{4}(2\theta^1)(-\frac{1}{(\theta^2)^2}) = \frac{1}{2}\frac{\theta^1}{(\theta^2)^2} = 2\mu\sigma^2$$

$$g_{22} = \frac{\partial^2\psi}{\partial\theta^2\partial\theta^2} = \frac{\partial}{\partial\theta^2}\left(\frac{(\theta^1)^2}{4(\theta^2)^2} + \frac{1}{2}\frac{1}{\theta^2}\right) = \frac{(\theta^1)^2}{4}\left(-2\frac{1}{(\theta^2)^3}\right) - \frac{1}{2}\frac{1}{(\theta^2)^2} = -\frac{1}{2(\theta^2)^2}\left[\frac{(\theta^1)^2}{\theta^2} + 1\right]$$

ここで，

$$\frac{1}{2(\theta^2)^2} = 2\sigma^4 \qquad \frac{(\theta^1)^2}{\theta^2} = -2\frac{\mu^2}{\sigma^2}$$

であるので，

$$g_{22} = -2\sigma^4 \left(-2\frac{\mu^2}{\sigma^2} + 1 \right) = 4\mu^2\sigma^2 - 2\sigma^4$$

となる．

$e-$ 測地線に沿った距離

ここで，2 つの正規分布の距離を計算してみる．スカラー正規分布多様体 M の 2 点 $p(x|\boldsymbol{\theta}_p) = \mathcal{N}(x|(\mu_p, \sigma_p))$ と $q(x|\boldsymbol{\theta}_q) = \mathcal{N}(x|(\mu_q, \sigma_q))$ を考え，p, q 間の距離を導出する．

$\boldsymbol{\theta} = (\theta^1, \theta^2)$ は

$$\theta^1 = \frac{\mu}{\sigma^2}$$
$$\theta^2 = -\frac{1}{2\sigma^2}$$

であるので，

$$\boldsymbol{\theta}_p = (\theta^1_p, \theta^2_p) = (\frac{\mu_p}{(\sigma_p)^2}, -\frac{1}{2(\sigma_p)^2}) \tag{4.33}$$

$$\boldsymbol{\theta}_q = (\theta^1_q, \theta^2_q) = (\frac{\mu_q}{(\sigma_q)^2}, -\frac{1}{2(\sigma_q)^2}) \tag{4.34}$$

である．ちなみに，p, q 間のユークリッド距離 D は,

$$D^2 = \|\boldsymbol{\theta}_p - \boldsymbol{\theta}_q\|^2 = \left(\frac{\mu_p}{(\sigma_p)^2} - \frac{\mu_q}{(\sigma_q)^2} \right)^2 + \left(\frac{1}{2(\sigma_p)^2} - \frac{1}{2(\sigma_q)^2} \right)^2$$

である．

多様体 M 上での p, q 間を結ぶ距離を求める．M は双対平坦な多様体で，座標 $\boldsymbol{\theta}$ は双対アフィン座標である．つまり，$\boldsymbol{\theta}$ 上で 2 点を結ぶ測地線は直線であり，パラメータ $t \in [0, 1]$ を用いて

$$\boldsymbol{\theta}(t) = (1-t)\boldsymbol{\theta}_p + t\boldsymbol{\theta}_q$$

と表される[7]．この直線は $e-$ 測地線と呼ばれる．各成分は

$$\theta^1(t) = (1-t)\theta^1_p + t\theta^1_q \qquad \theta^2(t) = (1-t)\theta^2_p + t\theta^2_q \tag{4.35}$$

と表される．2 点間の距離 L は

$$L = \int_0^1 \|\frac{d\boldsymbol{\theta}}{dt}\| dt$$

ここで,

$$\frac{d\boldsymbol{\theta}}{dt} = \boldsymbol{\theta}_q - \boldsymbol{\theta}_p \quad \text{であるので} \quad L = \int_0^1 \|\boldsymbol{\theta}_q - \boldsymbol{\theta}_p\| dt$$

---

[7]測地線が直線となるのは，測地線方程式

$$\frac{\partial^2 \theta^i}{\partial t^2} + \Gamma^i{}_{k\ell} \frac{d\theta^k}{dt} \frac{d\theta^\ell}{dt} = 0$$

において，双対アフィン座標 $\boldsymbol{\theta}$ では，接続係数 $\Gamma^i{}_{k\ell} = 0$ であるので，測地線方程式は $\partial^2\theta^i/\partial t^2 = 0$ となり，結局，測地線を表す式は $\boldsymbol{\theta} = \boldsymbol{a}t + \boldsymbol{b}$ と t に関する 1 次式で表される．

である．さらに，接ベクトルは

$$\boldsymbol{\theta}_q - \boldsymbol{\theta}_p = (\theta_q^1 - \theta_p^1)\frac{\partial}{\partial\theta^1} + (\theta_q^2 - \theta_p^2)\frac{\partial}{\partial\theta^2} = (\theta_q^i - \theta_p^i)\frac{\partial}{\partial\theta^i}$$

と書けるので，

$$\|\boldsymbol{\theta}_q - \boldsymbol{\theta}_p\| = \sqrt{g\left((\theta_q^i - \theta_p^i)\frac{\partial}{\partial\theta^i}, (\theta_q^j - \theta_p^j)\frac{\partial}{\partial\theta^j}\right)} = \sqrt{g_{11}(t)(\theta_q^1 - \theta_p^1)^2 + 2g_{12}(t)(\theta_q^1 - \theta_p^1)(\theta_q^2 - \theta_p^2) + g_{22}(t)(\theta_q^2 - \theta_p^2)^2}$$

である．ここで，

$$\begin{aligned} g_{11}(t) &= -\frac{1}{2\theta^2(t)} \\ g_{12}(t) &= \frac{1}{2}\frac{\theta^1(t)}{(\theta^2(t))^2} \\ g_{22}(t) &= -\frac{1}{2(\theta^2(t))^2}\left[\frac{(\theta^1(t))^2}{\theta^2(t)} + 1\right] \end{aligned}$$

であるので，結局

$$\begin{aligned} L = &\int_0^1 \sqrt{g_{11}(t)(\Delta^1)^2 + 2g_{12}(t)\Delta^1\Delta^2 + g_{22}(t)(\Delta^2)^2}tdt \\ &= \int_0^1 \sqrt{-(\Delta^1)^2\frac{1}{2\theta^2(t)} + \Delta^1\Delta^2\frac{\theta^1(t)}{(\theta^2(t))^2} - (\Delta^2)^2\frac{1}{2(\theta^2(t))^2}\left[\frac{(\theta^1(t))^2}{\theta^2(t)} + 1\right]}dt \end{aligned}$$

となる．ここで，

$$\Delta^1 = \theta_q^1 - \theta_p^1 = \frac{\mu_p}{(\sigma_p)^2} - \frac{\mu_q}{(\sigma_q)^2} \qquad \Delta^2 = \theta_q^2 - \theta_p^2 = -\frac{1}{2(\sigma_p)^2} + -\frac{1}{2(\sigma_q)^2}$$

である．

式 (4.35) や (4.33)，(4.34) を用いて，上式の積分を実行するのであるが.....

甘利 [3] によれば，ユークリッド空間における測地線である直線は１）接ベクトルが方向を変えない，２）2 点を結ぶ長さが最小である，の性質を有するが，双対平坦な空間における測地線である直線は１）の性質を有するが，２）の性質は（一般的には）持っていない．つまり，双対平坦な空間の 2 点間の最小距離を上の計算で求めることは出来ない（のであろう）．–接続を双対にするという “複雑さ” を受け入れたことで,（リーマン接続の “単純” さを犠牲にすることで,）平坦性という “単純” さを得たのではなかった!!

4.6 ダイバージェンス

4.6.1 Bregman ダイバージェンス

双対平坦な多様体 (M, ∇, ∇^*, g) を仮定し，∇ に関するアフィン座標を $\boldsymbol{\theta}$ とする．双対平坦な多様体においては，リーマン計量 g_{ij} が

$$g_{ij} = \frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}$$

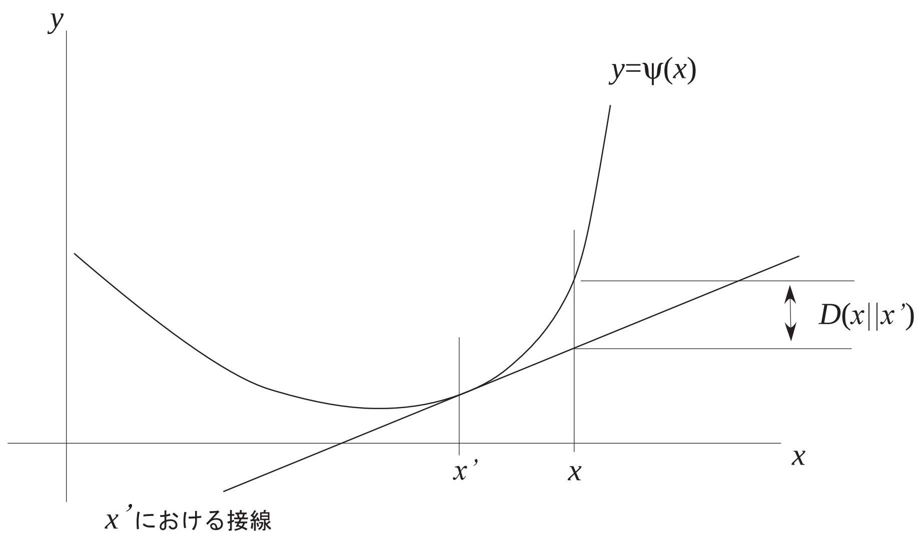

図 4.1: Bregman ダイバージェンスの説明．$\psi(x)$ を凸関数とする．x' で $\psi(x)$ に接する直線は $y=(x-x')\frac{\partial\psi}{\partial x}(x')+\psi(x')$ で与えられる．したがって，ダイバージェンス $D(x\|x')=\psi(x)-\left[(x-x')\frac{\partial\psi}{\partial x}(x')+\psi(x')\right]$ は，この接線と関数 $\psi(x)$ の x における乖離の大きさを表している．

と表されるような凸関数 $\psi(\boldsymbol{\theta})$ が必ず存在する．このとき M 上の 2 点 $p,q\in M$ に対して，

$$D(p\|q)=(\theta^i(q)-\theta^i(p))\frac{\partial\psi}{\partial\theta^i}(q)-(\psi(q)-\psi(p)) \tag{4.36}$$

を Bregman ダイバージェンスと呼ぶ．

Bregman ダイバージェンスは，明確な解釈を与えることができる．式 (4.36) を 1 次元で考えてみる．点 p,q における座標を x,x' とすれば，

$$D(x\|x')=(x'-x)\frac{d\psi}{dx}(x')-(\psi(x')-\psi(x))=\psi(x)-\left((x-x')\frac{d\psi}{dx}(x')+\psi(x')\right)$$

となる．図 4.1 に示すように，$\psi(x)$ は凸関数とし，x' で $\psi(x)$ に接する直線を引く．この直線は $y=(x-x')\dfrac{d\psi}{dx}(x')+\psi(x')$ で与えられる．したがって，ダイバージェンス $D(x\|x')=\psi(x)-\left[(x-x')\dfrac{d\psi}{dx}(x')+\psi(x')\right]$ は，この接線と関数 $\psi(x)$ の x における乖離の大きさを表している．この事から，（$\psi(x)$ は凸関数であるので）$D(x\|x')\geq 0$ は自明であり，等号成立は $x=x'$ の場合のみである．

一般の n 次元の場合においても，式 (4.36) によれば，点 q における凸関数 $\psi(\boldsymbol{\theta})$ の接平面の点 p における値 $(\theta^i(q)-\theta^i(p))\dfrac{\partial\psi}{\partial\theta^i}(q)-(\psi(q)$ と，点 p における凸関数の値 $\psi(p)$ の差が $D(p\|q)$ である．したがって，ダイバージェンスとは，双対平坦な多様体において，点 p,q 間の距離を，p において q における接平面とポテンシャル関数との乖離の大きさで評価する「偽距離」である．ポテンシャル関数は凸関数であるので，p,q 間の距離が大きくなれば乖離の大きさも大きくなる．

4.6.2　Bregman ダイバージェンスの双対アフィン座標での表現

Bregman ダイバージェンスは双対アフィン座標を用いて，

$$D(p\|q) = \psi(\theta(p)) + \varphi(\eta(q)) - \theta^i(p)\eta_i(q) \tag{4.37}$$

と表される．ここで，θ^i, η_i は M の双対アフィン座標系である．

Bregman ダイバージェンス（式 (4.36)）は式 (4.37) の定義と等価である．

証明： 証明は簡単である．上式のカッコを払えば，$\dfrac{\partial\psi}{\partial\theta^i}(q) = \eta_i(q)$ に注意して，

$$D(p\|q) = (\theta^i(q) - \theta^i(p))\eta_i(q) - (\psi(q) - \psi(p)) = \theta^i(q)\eta_i(q) - \theta^i(p)\eta_i(q) - \psi(q) + \psi(p)$$

であるが,

$$\theta^i(q)\eta_i(q) - \psi(q) = \varphi(q)$$

であるので,

$$D(p\|q) = \varphi(q) + \psi(p) - \theta^i(p)\eta_i(q)$$

となって，式 (4.37) を得る．

このダイバージェンスは $\alpha-$ ダイバージェンス，または $\nabla-$ ダイバージェンスなどと呼ばれる場合もある．ここでは，文献 [2] による最もシンプルな呼び方であるダイバージェンスを用いる．

4.6.3　ダイバージェンスの座標不変性

ダイバージェンスが双対アフィン座標系の取り方にはよらない事を示す．

定理：θ^i, η_i と $\widetilde{\theta}^i$ と $\widetilde{\eta}_i$ を M の 2 組の双対アフィン座標系とする．それぞれの双対ポテンシャルを $\psi(\theta), \varphi(\eta)$, $\widetilde{\psi}(\widetilde{\theta}), \widetilde{\varphi}(\widetilde{\eta})$ と書く．すると，

$$\psi(\theta(p)) + \varphi(\eta(q)) - \theta^i(p)\eta_i(q) = \widetilde{\psi}(\widetilde{\theta}(p)) + \widetilde{\varphi}(\widetilde{\eta}(q)) - \widetilde{\theta}^i(p)\widetilde{\eta}_i(q)$$

が成り立つ．

証明：アフィン座標系はあるアフィン変換で関係付けられている．

$$\widetilde{\theta}^\lambda = A_i^\lambda\theta^i + a^\lambda \qquad \widetilde{\eta}_\lambda = B_\lambda^i\eta_i + b_\lambda \tag{4.38}$$

$\boldsymbol{A}$ と $\boldsymbol{B}$ は正則行列である．$\widetilde{\partial}_\lambda = \dfrac{\partial}{\partial\widetilde{\theta}^\lambda}$ と書けば，ここで,

$$\frac{\partial}{\partial\theta^i} = \frac{\partial\widetilde{\theta}^\lambda}{\partial\theta^i}\frac{\partial}{\partial\widetilde{\theta}^\lambda} \quad \to \quad \partial_i = \frac{\partial\widetilde{\theta}^\lambda}{\partial\theta^i}\widetilde{\partial}_\lambda = A_i^\lambda\widetilde{\partial}_\lambda$$

であり，同様に

$$\frac{\partial}{\partial\eta_i} = \frac{\partial\widetilde{\eta}_\lambda}{\partial\eta_i}\frac{\partial}{\partial\widetilde{\eta}_\lambda} \quad \to \quad \partial^i = \frac{\partial\widetilde{\eta}_\lambda}{\partial\eta_i}\widetilde{\partial}^\lambda = B_\lambda^i\widetilde{\partial}^\lambda$$

である．

双対アフィン座標系においては $g(\partial_i, \partial^j) = g(\dfrac{\partial}{\partial\theta^i}, \dfrac{\partial}{\partial\eta_j}) = \delta_i^j$ の関係がある．したがって，

$$\delta_i^j = g(\partial_i, \partial^j) = g(A_i^\lambda \widetilde{\partial}_\lambda, B_\mu^j \widetilde{\partial}^\mu) = A_i^\lambda B_\mu^j g(\widetilde{\partial}_\lambda, \widetilde{\partial}^\mu) = A_i^\lambda B_\mu^j \delta_\lambda^\mu = A_i^\lambda B_\lambda^j$$

を得る．ここで，行列 $\boldsymbol{A}$ と $\boldsymbol{B}$ を

$$\boldsymbol{A} = \begin{bmatrix} A_1^1 & \cdots & A_1^n \\ \vdots & \ddots & \vdots \\ A_n^1 & \cdots & A_n^n \end{bmatrix} \qquad \boldsymbol{B} = \begin{bmatrix} B_1^1 & \cdots & B_1^n \\ \vdots & \ddots & \vdots \\ B_n^1 & \cdots & B_n^n \end{bmatrix}$$

とすれば，すなわち，上付きの添え字が列を，下付きが行を表すように行列を定義すれば

$$A_i^\lambda B_\lambda^j = \delta_i^j$$

は

$$\boldsymbol{AB} = \boldsymbol{I}$$

を表す．先ほど導いた式 $\partial_i = A_i^\lambda \widetilde{\partial}_\lambda$ から

$$\partial_i = A_i^\lambda \widetilde{\partial}_\lambda \quad \to \quad \begin{bmatrix} \partial_1 \\ \vdots \\ \partial_n \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} \widetilde{\partial}_1 \\ \vdots \\ \widetilde{\partial}_n \end{bmatrix} \quad \to \quad \begin{bmatrix} \widetilde{\partial}_1 \\ \vdots \\ \widetilde{\partial}_n \end{bmatrix} = \boldsymbol{B} \begin{bmatrix} \partial_1 \\ \vdots \\ \partial_n \end{bmatrix} \quad \to \quad \widetilde{\partial}_\lambda = B_\lambda^i \partial_i$$

を得る．

全く同様に

$$\partial^i = B_\lambda^i \widetilde{\partial}^\lambda$$

であるので，

$$[\partial^1, \ldots, \partial^n] = [\widetilde{\partial}^1, \ldots, \widetilde{\partial}^n]\boldsymbol{B} = [\partial^1, \ldots, \partial^n] = [\widetilde{\partial}^1, \ldots, \widetilde{\partial}^n] \begin{bmatrix} B_1^1 & \cdots & B_1^n \\ \vdots & \ddots & \vdots \\ B_n^1 & \cdots & B_n^n \end{bmatrix}$$

が成り立つので，

$$[\widetilde{\partial}^1, \ldots, \widetilde{\partial}^n] = [\partial^1, \ldots, \partial^n]\boldsymbol{A} = [\partial^1, \ldots, \partial^n] \begin{bmatrix} A_1^1 & \cdots & A_1^n \\ \vdots & \ddots & \vdots \\ A_n^1 & \cdots & A_n^n \end{bmatrix}$$

が成り立つ．つまり，

$$\widetilde{\partial}^\lambda = A_i^\lambda \partial^i$$

を得る．

いよいよ証明に入る．まず，ψ と $\widetilde{\psi}$ の関係を導く．$\eta_i = \dfrac{\partial\psi}{\partial\theta^i}$ から，

$$\widetilde{\eta}_\lambda = \frac{\partial\widetilde{\psi}}{\partial\widetilde{\theta}^\lambda} = \frac{\partial}{\partial\widetilde{\theta}^\lambda}\widetilde{\psi} = B_\lambda^j \frac{\partial}{\partial\theta^j}\widetilde{\psi}$$

一方，

$$\widetilde{\eta}_\lambda = B_\lambda^j \eta_j + b_\lambda = B_\lambda^j \frac{\partial\psi}{\partial\theta^j} + b_\lambda$$

上2つの式を見比べて，

$$B^j_\lambda \frac{\partial}{\partial \theta^j}\widetilde{\psi} = B^j_\lambda \frac{\partial}{\partial \theta^j}\psi + b_\lambda$$

を得る．これは，行列と列ベクトルで書くと，

$$\boldsymbol{B}\begin{bmatrix} \dfrac{\partial \widetilde{\psi}}{\partial \theta^1} \\ \vdots \\ \dfrac{\partial \widetilde{\psi}}{\partial \theta^n} \end{bmatrix} = \boldsymbol{B}\begin{bmatrix} \dfrac{\partial \psi}{\partial \theta^1} \\ \vdots \\ \dfrac{\partial \psi}{\partial \theta^n} \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

である．したがって，

$$\begin{bmatrix} \dfrac{\partial \widetilde{\psi}}{\partial \theta^1} \\ \vdots \\ \dfrac{\partial \widetilde{\psi}}{\partial \theta^n} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial \psi}{\partial \theta^1} \\ \vdots \\ \dfrac{\partial \psi}{\partial \theta^n} \end{bmatrix} + \boldsymbol{A}\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

が成り立ち，以下の式を得る．

$$\frac{\partial}{\partial \theta^j}\widetilde{\psi} = \frac{\partial}{\partial \theta^j}\psi + [\boldsymbol{B}^{-1}]^\lambda_j b_\lambda \quad \to \quad \frac{\partial}{\partial \theta^j}\widetilde{\psi} = \frac{\partial}{\partial \theta^j}\psi + A^\lambda_j b_\lambda$$

さらに，両辺を θ^j で積分して，

$$\widetilde{\psi}(\widetilde{\theta}) = \psi(\theta) + A^\lambda_j b_\lambda \theta^j + C \tag{4.39}$$

を得る．ここで，C は積分定数である，

　次に，φ と $\widetilde{\varphi}$ の関係を導く．

$$\psi(\theta) + \varphi(\eta) = \theta^k \eta_k$$

を用いれば，

$$\begin{aligned}\widetilde{\varphi}(\widetilde{\eta}) = \widetilde{\theta}^k\widetilde{\eta}_k - \widetilde{\psi}(\widetilde{\theta}) &= (A^k_i\theta^i + a^k)(B^j_k\eta_j + b_k) - (\psi(\theta) + A^\lambda_j b_\lambda\theta^j + C) \\ &= \theta^i\eta_i - \psi(\theta) + a^kB^j_k\eta_j + a^kb_k - C = \varphi(\eta) + a^kB^j_k\eta_j + a^kb_k - C \end{aligned} \tag{4.40}$$

を得る．したがって，式 (4.38), (4.39), (4.40) を用いれば，

$$\begin{aligned}&\widetilde{\psi}(\widetilde{\theta}(p)) + \widetilde{\varphi}(\widetilde{\eta}(q)) - \widetilde{\theta}^k(p)\widetilde{\eta}_k(q) \\ &\quad = \big(\psi(\theta(p)) + A^\lambda_j b_\lambda\theta^j(p) + C\big) + \Big(\varphi(\eta(q)) + a^kB^j_k\eta_j(q) + a^kb_k - C\Big) - (A^k_i\theta^i(p) + a^k)(B^j_k\eta_j(q) + b_k) \\ &\quad = \psi(\theta(p)) + \varphi(\eta(q)) - \theta^i(p)\eta_i(q)\end{aligned} \tag{4.41}$$

を得る．すなわち，ダイバージェンスは座標変換に対して不変であり，幾何学的に固有な量である．

4.6.4　ダイバージェンスの性質

　$p \in M$ の θ 座標と η 座標を

$$\theta(p) = (\theta^1(p), \ldots, \theta^n(p)) \qquad \eta(p) = (\eta_1(p), \ldots, \eta_n(p))$$

と表せば，式 (3.21) と (3.22)：

$$\varphi(\eta(p)) = \max_{q \in M} \left(\theta^i(q)\eta_i(p) - \psi(\theta(q)) \right)$$
$$\psi(\theta(p)) = \max_{q \in M} \left(\eta_i(q)\theta^i(p) - \varphi(\eta(q)) \right)$$

が成り立つ．この事から，$p \neq q$ の場合では，

$$\varphi(\eta(p)) \geq \theta^i(q)\eta_i(p) - \psi(\theta(q)) \to \varphi(\eta(p)) + \psi(\theta(q)) - \theta^i(q)\eta_i(p) \geq 0$$

および

$$\psi(\theta(q)) \geq \eta_i(p)\theta^i(q) - \varphi(\eta(p)) \to \psi(\theta(q)) + \varphi(\eta(p)) - \eta_i(p)\theta^i(q) \geq 0$$

であり，

$$D(p\|q) \geq 0$$

を得る．また，等号成立は $p = q$ の場合のみである．ただし，この事は図 4.1 からほとんど自明である．

4.6.5 ユークリッド空間での例

$\nabla^* = \nabla$ である場合は自己双対と呼ばれる．これは，リーマン接続をもつリーマン多様体の場合である．この場合，双対平坦性は，単なる平坦性であり，平坦なリーマン多様体はユークリッド空間に帰着する．

$$\psi(\theta) + \varphi(\eta) = \theta^k \eta_k$$

において，$\theta^i = \eta_i = z^i$ として，$\psi(z) = \varphi(z)$ とすれば，

$$2\psi(z) = \sum_{i=1}^{n} (z^i)^2$$

すなわち，

$$\psi(z) = \varphi(z) = \frac{1}{2} \sum_{i=1}^{n} (z^i)^2$$

を得る．したがって，

$$D(p\|q) = \psi(z(p)) + \psi(z(q)) - \sum_{i=1}^{n} z^i(p) z^i(q) = \frac{1}{2} \sum_{i=1}^{n} (z^i(p))^2 + \frac{1}{2} \sum_{i=1}^{n} (z^i(q))^2 - \sum_{i=1}^{n} z^i(p) z^i(q) = \frac{1}{2} \sum_{i=1}^{n} (z^i(p) - z^i(q))^2$$

となる．

ダイバージェンスは距離の 2 乗の次元を持っているが，ユークリッド距離とは異なる（$\frac{1}{2}$ がついている.）．さらに，距離の公理のうちで，対称性や三角不等式はなりたたない．

4.6.6 一般化したピタゴラスの定理

一般化したピタゴラスの定理が成り立つ．

双対平坦多様体 (M, g, ∇, ∇^*) 上に 3 点 p, q, r を取る．もし，p と q を結ぶ $\nabla-$ 測地線（接続 ∇ から計算される測地線）が，q と r を結ぶ ∇^*- 測地線（接続 ∇^* から計算される測地線）が q において直交するなら

$$D(p\|q) + D(q\|r) = D(p\|r)$$

が成り立つ．

証明： アフィン座標系に対する接続係数はゼロなので，測地線は $\dfrac{d^2x^i}{dt^2} = 0$ の解であり，直線になる．したがって，p, q を結ぶ測地線は,

$$\boldsymbol{\theta}(t) = (\theta^1(t), \ldots, \theta^n(t)) = t\boldsymbol{\theta}(p) + (1-t)\boldsymbol{\theta}(q)$$

と表され，q, r を結ぶ測地線は

$$\boldsymbol{\eta}(t) = (\eta_1(t), \ldots, \eta_n(t)) = t\boldsymbol{\eta}(r) + (1-t)\boldsymbol{\eta}(q))$$

と表される．この 2 つの測地線が q において直交するので,

$$g\left(\frac{d\boldsymbol{\theta}(t)}{dt}(0), \frac{d\boldsymbol{\eta}(t)}{dt}(0)\right) = 0$$

が成り立つ．ところで,

$$\frac{d\boldsymbol{\theta}(t)}{dt} = \frac{\partial\theta^i(t)}{\partial t}\frac{\partial}{\partial\theta^i} = (\theta^i(p) - \theta^i(q))\frac{\partial}{\partial\theta^i} \qquad \frac{d\boldsymbol{\eta}(t)}{dt} = \frac{\partial\eta_i(t)}{\partial t}\frac{\partial}{\partial\eta_i} = (\eta_i(r) - \eta_i(q))\frac{\partial}{\partial\eta_i}$$

である[8]．したがって,

$$\begin{aligned} g\left(\frac{d\boldsymbol{\theta}(t)}{dt}, \frac{d\boldsymbol{\eta}(t)}{dt}\right) &= g\left((\theta^i(p) - \theta^i(q))\frac{\partial}{\partial\theta^i}, (\eta_j(r) - \eta_j(q))\frac{\partial}{\partial\eta_j}\right) \\ &= (\theta^i(p) - \theta^i(q))(\eta_j(r) - \eta_j(q))\delta^i_j = \Big(\theta^i(p) - \theta^i(q)\Big)\Big(\eta_i(r) - \eta_i(q)\Big) = 0 \end{aligned} \tag{4.42}$$

が成り立つ．また，ダイバージェンスの定義（式 (4.37)）より,

$$\begin{aligned} D(p\|q) &= \psi(\theta(p)) + \varphi(\eta(q)) - \theta^i(p)\eta_i(q) \\ D(q\|r) &= \psi(\theta(q)) + \varphi(\eta(r)) - \theta^i(q)\eta_i(r) \\ D(p\|r) &= \psi(\theta(p)) + \varphi(\eta(r)) - \theta^i(p)\eta_i(r) \end{aligned}$$

であるので,

$$\begin{aligned} &D(p\|q) + D(q\|r) - D(p\|r) \\ &\quad = \psi(\theta(p)) + \varphi(\eta(q)) - \theta^i(p)\eta_i(q) + \psi(\theta(q)) + \varphi(\eta(r)) - \theta^i(q)\eta_i(r) - \Big[\psi(\theta(p)) + \varphi(\eta(r)) - \theta^i(p)\eta_i(r)\Big] \\ &= \psi(\theta(q)) + \varphi(\eta(q)) - \theta^i(p)\eta_i(q) - \theta^i(q)\eta_i(r) + \theta^i(p)\eta_i(r) = \theta^i(q)\eta_i(q) - \theta^i(p)\eta_i(q) - \theta^i(q)\eta_i(r) + \theta^i(p)\eta_i(r) \\ &= \Big(\theta^i(p) - \theta^i(q)\Big)\Big(\eta_i(r) - \eta_i(q)\Big) = 0 \end{aligned}$$

となって,

$$D(p\|q) + D(q\|r) = D(p\|r)$$

を示すことができた．

[8]注意：$(\theta^i(p) - \theta^i(q))\dfrac{\partial}{\partial\theta^i}$ は接ベクトルの基底ベクトル表示である．($\dfrac{\partial}{\partial\theta^i}$ は x^i 方向の基底ベクトルである．)

4.6.7　指数型分布族：KL ダイバージェンスの導出

Bregman ダイバージェンスは，指数型分布族の確率分布

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \exp[c(\boldsymbol{x}) + \theta^i u_i(\boldsymbol{x}) - \psi(\boldsymbol{\theta})] \tag{4.43}$$

を仮定すれば KL ダイバージェンスに等しくなる．以下これを示す．この節では，指数分布族の統計多様体 M に対して，$\alpha = \pm 1$ のアルファ接続 $\nabla^{(\alpha)}, \nabla^{(-\alpha)}$ を仮定する．

ここで，

$$\eta_i = E_{\boldsymbol{\theta}}\left[u_i(\boldsymbol{x})\right]$$

とすれば，θ^i と η_i は互いに双対なアフィン局所座標系となっている．ポテンシャル $\varphi(\boldsymbol{\eta})$ は，式 (4.31) から

$$\varphi(\boldsymbol{\eta}) = \theta^i \eta_i - \psi(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}[\log p(\boldsymbol{x}, |\theta) - c(\boldsymbol{x})]$$

と求まる．上式は $p, q \in M$ で成り立つので

$$\begin{aligned}\varphi(\boldsymbol{\eta}(p)) &= E_{\boldsymbol{\theta}(p)}[\log p(\boldsymbol{x}, |\theta(p)) - c(\boldsymbol{x})] \\ \varphi(\boldsymbol{\eta}(q)) &= E_{\boldsymbol{\theta}(q)}[\log q(\boldsymbol{x}, |\theta(q)) - c(\boldsymbol{x})]\end{aligned}$$

である．

M 上に 2 点 p, q を取り，ダイバージェンスを計算すれば

$$D(p\|q) = \psi(\boldsymbol{\theta}(p)) + \varphi(\boldsymbol{\eta}(q)) - \theta^i(p)\eta_i(q) = \psi(\boldsymbol{\theta}(p)) + E_{\boldsymbol{\theta}(q)}[\log q(\boldsymbol{x}, |\theta(q)) - c(\boldsymbol{x})] - \theta^i(p) E_{\boldsymbol{\theta}(q)}[u_i(\boldsymbol{x})]$$

となる．ところで，点 p における指数型分布族の確率分布にもどると

$$p(\boldsymbol{x}|\boldsymbol{\theta}(p)) = \exp[c(\boldsymbol{x}) + \theta^i(p) u_i(\boldsymbol{x}) - \psi(\boldsymbol{\theta}(p))]$$

である．上式の log を取りその両辺の，確率分布 $q(\boldsymbol{x}|\boldsymbol{\theta}(q))$ に関する期待値をとれば，

$$E_{\boldsymbol{\theta}(q)}[\log p(\boldsymbol{x}|\boldsymbol{\theta}(p)) - c(\boldsymbol{x})] = \theta^i(p) E_{\boldsymbol{\theta}(q)}[u_i(\boldsymbol{x})] - \psi(\boldsymbol{\theta}(p))$$

を得る．つまり

$$\psi(\boldsymbol{\theta}(p)) - \theta^i(p) E_{\boldsymbol{\theta}(q)}[u_i(\boldsymbol{x})] = -E_{\boldsymbol{\theta}(q)}[\log p(\boldsymbol{x}|\boldsymbol{\theta}(p)) - c(\boldsymbol{x})]$$

が成り立つのでダイバージェンス $D(p\|q)$ は

$$\begin{aligned}D(p\|q) &= E_{\boldsymbol{\theta}(q)}[\log q(\boldsymbol{x}, |\theta(q)) - c(\boldsymbol{x})] - E_{\boldsymbol{\theta}(q)}[\log p(\boldsymbol{x}|\boldsymbol{\theta}(p)) - c(\boldsymbol{x})] \\ &= E_{\boldsymbol{\theta}(q)}[\log q(\boldsymbol{x}, |\theta(q)) - c(\boldsymbol{x}) - \log p(\boldsymbol{x}|\boldsymbol{\theta}(p)) + c(\boldsymbol{x})] = E_{\boldsymbol{\theta}(q)}[\log q(\boldsymbol{x}, |\theta(q)) - \log p(\boldsymbol{x}|\boldsymbol{\theta}(p))] \\ &= \sum_{\boldsymbol{x} \in \Omega} q(\boldsymbol{x}|\boldsymbol{\theta}(q)) \log \frac{q(\boldsymbol{x}|\boldsymbol{\theta}(q))}{p(\boldsymbol{x}|\boldsymbol{\theta}(p))}\end{aligned} \tag{4.44}$$

となる．これは KL ダイバージェンスに等しい．

4.7 Appendix: いくつかの証明

4.7.1 平坦な多様体

定理 アフィン接続 ∇ を持つ多様体 M において，次の 2 条件は同値である．

(i) M は平坦である．

(ii) M の各点の周りにアフィン座標系が存在する．

証明： (i)→(ii)

まず，接続係数の変換測を導く．

局所座標系 x^i において[9]，x^j 方向の接ベクトルの x^i 方向への方向微分は,

$$\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = \Gamma^k{}_{ij} \frac{\partial}{\partial x^k}$$

と書ける．また，別の局所座標系 ξ^a を用いると，

$$\nabla_{\frac{\partial}{\partial \xi^a}} \frac{\partial}{\partial \xi^b} = \Gamma^c{}_{ab} \frac{\partial}{\partial \xi^c}$$

と書ける．2 つの座標系で表すことのできる領域では，接続係数 $\Gamma^k{}_{ij}$ と $\Gamma^c{}_{ab}$ の間に何らかの変換測があるはずである．この変換測を導く．

そのため，$\nabla_{\frac{\partial}{\partial x^i}} \dfrac{\partial}{\partial x^j}$ を 2 通りで計算する．まず，

$$\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = \Gamma^k{}_{ij} \frac{\partial}{\partial x^k} = \Gamma^k{}_{ij} \frac{\partial \xi^c}{\partial x^k} \frac{\partial}{\partial \xi^c} \tag{4.45}$$

と表せる．上式は，接ベクトル変換測：$\dfrac{\partial}{\partial x^k} = \dfrac{\partial \xi^c}{\partial x^k} \dfrac{\partial}{\partial \xi^c}$ を用いただけである[10]．

一方，この変換測を用いて，

$$\begin{aligned}
\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} &= \nabla_{\frac{\partial}{\partial x^i}} \left(\frac{\partial \xi^b}{\partial x^j} \frac{\partial}{\partial \xi^b} \right) = \frac{\partial^2 \xi^b}{\partial x^i \partial x^j} \frac{\partial}{\partial \xi^b} + \frac{\partial \xi^b}{\partial x^j} \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial \xi^b} = \frac{\partial^2 \xi^b}{\partial x^i \partial x^j} \frac{\partial}{\partial \xi^b} + \frac{\partial \xi^b}{\partial x^j} \nabla_{\frac{\partial \xi^a}{\partial x^i} \frac{\partial}{\partial \xi^a}} \frac{\partial}{\partial \xi^b} \\
&= \frac{\partial^2 \xi^b}{\partial x^i \partial x^j} \frac{\partial}{\partial \xi^b} + \frac{\partial \xi^b}{\partial x^j} \frac{\partial \xi^a}{\partial x^i} \nabla_{\frac{\partial}{\partial \xi^a}} \frac{\partial}{\partial \xi^b} = \frac{\partial^2 \xi^b}{\partial x^i \partial x^j} \frac{\partial}{\partial \xi^b} + \frac{\partial \xi^b}{\partial x^j} \frac{\partial \xi^a}{\partial x^i} \Gamma^c{}_{ab} \frac{\partial}{\partial \xi^c} \\
&= \left(\frac{\partial^2 \xi^c}{\partial x^i \partial x^j} + \frac{\partial \xi^b}{\partial x^j} \frac{\partial \xi^a}{\partial x^i} \Gamma^c{}_{ab} \right) \frac{\partial}{\partial \xi^c}
\end{aligned} \tag{4.46}$$

[9] 局所座標系 $x^1, x^2, \ldots, x^n$ と書くべきところを省略してこのように書いた．

[10] この変換測は，より明示的に書くと，

$$\begin{bmatrix} \frac{\partial}{\partial x^1} \\ \vdots \\ \frac{\partial}{\partial x^n} \end{bmatrix} = \begin{bmatrix} \frac{\partial \xi^1}{\partial x^1} & \cdots & \frac{\partial \xi^n}{\partial x^1} \\ \vdots & \ddots & \vdots \\ \frac{\partial \xi^1}{\partial x^n} & \cdots & \frac{\partial \xi^n}{\partial x^n} \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial \xi^1} \\ \vdots \\ \frac{\partial}{\partial \xi^n} \end{bmatrix}$$

となる．接ベクトルにヤコビ行列をかけて別の座標系の接ベクトルを得ている．

である．ここで，式 (4.45) と (4.46) を比較すれば，

$$\Gamma^k{}_{ij}\,\frac{\partial \xi^c}{\partial x^k} = \frac{\partial^2 \xi^c}{\partial x^i \partial x^j} + \frac{\partial \xi^b}{\partial x^j}\frac{\partial \xi^a}{\partial x^i}\Gamma^c{}_{ab}$$

を得る．ヤコビ行列 $\dfrac{\partial \xi^c}{\partial x^k}$ の逆行列を両辺に乗じて

$$\Gamma^k{}_{ij} = \frac{\partial^2 \xi^c}{\partial x^i \partial x^j}\frac{\partial x^k}{\partial \xi^c} + \frac{\partial \xi^b}{\partial x^j}\frac{\partial \xi^a}{\partial x^i}\frac{\partial x^k}{\partial \xi^c}\Gamma^c{}_{ab}$$

を得る．上式が，座標変換 $(\xi^1,\ldots,\xi^n)\to(x^1,\ldots,x^n)$ に関する接続係数の座標変換測である．

ここから，(i)→(ii) の証明に入る． 任意の局所座標系 $x^1,\ldots,x^n$ に対して，ある座標系 $\xi^1,\ldots,\xi^n$ が存在して，その座標系における接続係数はゼロとなるので，

$$0 = \Gamma^c{}_{ab} = \frac{\partial^2 x^\ell}{\partial \xi^a \partial \xi^b}\frac{\partial \xi^c}{\partial x^\ell} + \frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b}\frac{\partial \xi^c}{\partial x^k}\Gamma^k{}_{ij} \tag{4.47}$$

となる．以下では，このような座標系 (ξ^a) が，任意の座標系 (x^i) に対して必ず存在することを示す．このため，まず，恒等式

$$\frac{\partial x^\ell}{\partial \xi^b}\frac{\partial \xi^b}{\partial x^j} = \delta^\ell_j$$

の両辺を x^i で偏微分する．

$$\frac{\partial^2 x^\ell}{\partial \xi^a \partial \xi^b}\frac{\partial \xi^a}{\partial x^i}\frac{\partial \xi^b}{\partial x^j} + \frac{\partial x^\ell}{\partial \xi^b}\frac{\partial^2 \xi^b}{\partial x^i \partial x^j} = 0$$

であるので，両辺にヤコビ行列 $\dfrac{\partial \xi^a}{\partial x^i}$ と $\dfrac{\partial \xi^b}{\partial x^j}$ の逆行列をかけると，

$$\frac{\partial^2 x^\ell}{\partial \xi^a \partial \xi^b} = -\frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b}\frac{\partial x^\ell}{\partial \xi^d}\frac{\partial^2 \xi^d}{\partial x^i \partial x^j}$$

を得る．式 (4.47) に代入して，

$$\begin{aligned}
0 = \Gamma^c{}_{ab} &= \frac{\partial^2 x^\ell}{\partial \xi^a \partial \xi^b}\frac{\partial \xi^c}{\partial x^\ell} + \frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b}\frac{\partial \xi^c}{\partial x^k}\Gamma^k{}_{ij} = -\frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b}\frac{\partial x^\ell}{\partial \xi^d}\frac{\partial^2 \xi^d}{\partial x^i \partial x^j}\frac{\partial \xi^c}{\partial x^\ell} + \frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b}\frac{\partial \xi^c}{\partial x^k}\Gamma^k{}_{ij} \\
&= \frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b}\frac{\partial \xi^c}{\partial x^k}\Gamma^k{}_{ij} - \frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b}\frac{\partial^2 \xi^c}{\partial x^i \partial x^j} = \frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b}\left(\frac{\partial \xi^c}{\partial x^k}\Gamma^k{}_{ij} - \frac{\partial^2 \xi^c}{\partial x^i \partial x^j}\right)
\end{aligned} \tag{4.48}$$

を得る．したがって，$\Gamma^c{}_{ab} = 0$ が成り立つのは，

$$\frac{\partial \xi^c}{\partial x^k}\Gamma^k{}_{ij} = \frac{\partial^2 \xi^c}{\partial x^i \partial x^j} \tag{4.49}$$

が成り立つのと同値である．上式は $\xi^1,\ldots,\xi^n$ に関する 2 階の偏微分方程式である．この方程式が解 $\xi^1,\ldots,\xi^n$ を持てばこの座標系における $\Gamma^c{}_{ab}$ は恒等的にゼロとなる．

以下では，条件 (i)，つまり曲率テンソル $R=0$ と捩率テンソル $T=0$ の条件下で，式 (4.49) が（任意の座標系 (x^i) に対して）必ず解を持つことを示す．これは，偏微分方程式の解の存在条件に関する議論である．

式 (4.49) を，2 つに分けて書くと

$$\frac{\partial \xi^c}{\partial x^k} = \theta^c_k$$

$$\frac{\partial \theta^c_j}{\partial x^i} = \theta^c_k \Gamma^k{}_{ij}$$

との，連立微分方程式となる．この連立微分方程式が解を持つ条件（可積分条件）は

$$\frac{\partial^2 \xi^c}{\partial x^i \partial x^j} = \frac{\partial^2 \xi^c}{\partial x^j \partial x^i} \tag{4.50}$$

$$\frac{\partial^2 \theta^c_k}{\partial x^i \partial x^j} = \frac{\partial^2 \theta^c_k}{\partial x^j \partial x^i} \tag{4.51}$$

である．まず，式 (4.50) は

$$\frac{\partial^2 \xi^c}{\partial x^i \partial x^j} = \frac{\partial^2 \xi^c}{\partial x^j \partial x^i} \iff \frac{\partial}{\partial x^i}\theta^c_j = \frac{\partial}{\partial x^j}\theta^c_i \iff \theta^c_k {\Gamma^k}_{ij} = \theta^c_k {\Gamma^k}_{ji} \iff \theta^c_k T^k_{ij} = 0$$

となる．ここで，捩率テンソル $T^k_{ij} = 0$ であるので，この可積分条件は成り立つ．次に，式 (4.51) は，

$$\begin{aligned}
\frac{\partial^2 \theta^c_k}{\partial x^i \partial x^j} = \frac{\partial^2 \theta^c_k}{\partial x^j \partial x^i} &\iff \frac{\partial}{\partial x^i}(\theta^c_\ell {\Gamma^\ell}_{jk}) = \frac{\partial}{\partial x^j}(\theta^c_\ell {\Gamma^\ell}_{ik}) \\
&\iff \frac{\partial \theta^c_\ell}{\partial x^i}{\Gamma^\ell}_{jk} + \theta^c_\ell \frac{\partial {\Gamma^\ell}_{jk}}{\partial x^i} - \frac{\partial \theta^c_\ell}{\partial x^j}{\Gamma^\ell}_{ik} - \theta^c_\ell \frac{\partial {\Gamma^\ell}_{ik}}{\partial x^j} = 0 \\
&\iff \theta^c_m {\Gamma^m}_{i\ell} {\Gamma^\ell}_{jk} + \theta^c_m \frac{\partial {\Gamma^m}_{jk}}{\partial x^i} - \theta^c_m {\Gamma^m}_{j\ell} {\Gamma^\ell}_{ik} - \theta^c_m \frac{\partial {\Gamma^m}_{ik}}{\partial x^j} = 0 \\
&\iff \theta^c_m \left({\Gamma^m}_{i\ell} {\Gamma^\ell}_{jk} + \frac{\partial {\Gamma^m}_{jk}}{\partial x^i} - {\Gamma^m}_{j\ell} {\Gamma^\ell}_{ik} - \frac{\partial {\Gamma^m}_{ik}}{\partial x^j} \right) = 0 \iff \theta^c_m R^m_{ijk} = 0
\end{aligned}$$

となる．リーマン曲率テンソル $R^m_{ijk} = 0$ であるので，この可積分条件も成り立つ．したがって，式 (4.49) に示す微分方程式は解 $\xi^1, \ldots, \xi^n$ を持ち，この解のもとで接続係数は全てゼロとなる．

4.7.2　統計多様体に関する証明

∇ と ∇^* を互いに双対な接続とする．次の 4 条件のうち，任意の 2 条件を仮定すると，残りの 2 条件が成り立つ．

(1) ∇ の捩率はゼロである．

(2) ∇^* の捩率はゼロである．

(3) $(\nabla g)(X, Y, Z)$ は X, Y, Z に関して対称である．(コダッチの方程式)

(4) $\bar{\nabla} = \frac{1}{2}(\nabla + \nabla^*)$ は Riemann 接続である．

ここで，

$$(\nabla g)(X, Y, Z) = Xg(Y, Z) - g(\nabla_X Y, Z) - g(Y, \nabla_X Z) \tag{4.52}$$

である．これは接続 ∇ が「どのくらい計量的でないか」を測るテンソル量である．

証明：　まず，式 (4.52) を書き直す．

$$(\nabla g)(X, Y, Z) = g(\nabla^*_X Y, Z) - g(\nabla_X Y, Z)$$

と書き直せる．（この式の導出は未確認.）X と Y を入れ替えて，

$$(\nabla g)(Y, X, Z) = g(\nabla^*_Y X, Z) - g(\nabla_Y X, Z)$$

であるので，

$$(\nabla g)(X, Y, Z) - (\nabla g)(Y, X, Z) = g(\nabla^*_X Y - \nabla^*_Y X, Z) - g(\nabla_X Y - \nabla_Y X, Z)$$

を得る．また，捩率について，

$$T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y]$$
$$T^*(X,Y) = \nabla^*_X Y - \nabla^*_Y X - [X,Y]$$

であるので，結局，

$$\begin{aligned}&(\nabla g)(X,Y,Z) - (\nabla g)(Y,X,Z)\\&\quad = g(T^*(X,Y) + [X,Y], Z) - g(T(X,Y) + [X,Y], Z) = g(T^*(X,Y) - T(X,Y), Z) \quad (4.53)\end{aligned}$$

である．

(1),(2) が成り立っている場合：

式 (4.53) より，

$$(\nabla g)(X,Y,Z) = (\nabla g)(Y,X,Z)$$

が成り立つ．Y, Z の対称性は自明であるため，(3) が成り立つ．また，$T = T^* = 0$ が仮定されているため，$\bar{\nabla}$ は Riemann 接続である．したがって，(3),(4) が言える．

(1),(3) を仮定した場合：

式 (4.53) より，$T = T^*$ が言えるので，$T^* = 0$ が言える．上と同じ論法で $\bar{\nabla}$ は Riemann 接続である．(2),(3) を仮定した場合も同じである．

(1),(4) を仮定した場合：

$\nabla^* = 2\bar{\nabla} - \nabla$ を

$$T^*(X,Y) = \nabla^*_X Y - \nabla^*_Y X - [X,Y]$$

に代入すれば，

$$T^*(X,Y) = (2\bar{\nabla}_X - \nabla_X)Y - (2\bar{\nabla}_Y - \nabla_Y)X - [X,Y] = 2\bar{T}(X,Y) - T(X,Y) = 0$$

なぜなら，$\bar{\nabla}$ は Riemann 接続なので，$\bar{T} = 0$．したがって，(2) が言える．(1),(2) から (3) が言える．(2),(4) を仮定した場合も同じである．

(3),(4) を仮定した場合：

式 (4.53) より，

$$g(T^*(X,Y) - T(X,Y), Z) = 0$$

Z は任意なので，

$$T^*(X,Y) - T(X,Y) = 0$$

(4) より，$\bar{\nabla}$ はリーマン接続なので，その捩率

$$\frac{1}{2}(T(X,Y) + T^*(X,Y)) = 0 \quad \to \quad T(X,Y) + T^*(X,Y) = 0$$

したがって，$T = T^* = 0$ が言える．

文献 [2] における (1),(2)→(3) の証明： まず最初に以下の成立を示す．

$$(\nabla g)(X,Y,Z) + g(Y, T(X,Z)) = (\nabla g)(Z,Y,X) + g(Y, T^*(X,Z))$$

（文献 [2] では，$(\nabla g)(X,Y,Z)$, $(\nabla g)(Z,Y,X)$ はそれぞれ $(\nabla_X g)(Y,Z)$, $(\nabla_Z g)(Y,X)$ と表されている.）

証明：

$$\begin{aligned}(\nabla g)(X,Y,Z)+g(Y,T(X,Z)) &= Xg(Y,Z)-g(\nabla_X Y,Z)-g(Y,\nabla_X Z)+g(Y,\nabla_X Z-\nabla_Z X-[X,Z])\\ &= Xg(Y,Z)-g(\nabla_X Y,Z)-g(Y,\nabla_X Z)+g(Y,\nabla_X Z)-g(Y,\nabla_Z X)-g(Y,[X,Z])\\ &= Xg(Y,Z)-g(\nabla_X Y,Z)-g(Y,\nabla_Z X)-g(Y,[X,Z])\end{aligned}$$

また,

$$\begin{aligned}(\nabla g)(Z,Y,X)+g(Y,T^*(X,Z)) &= Zg(Y,X)-g(\nabla_Z Y,X)-g(Y,\nabla_Z X)+g(Y,\nabla^*_X Z-\nabla^*_Z X-[X,Z])\\ &= Zg(Y,X)-g(\nabla_Z Y,X)-g(Y,\nabla_Z X)+g(Y,\nabla^*_X Z)-g(Y,\nabla^*_Z X)-g(Y,[X,Z])\end{aligned}$$

である．ここで，双対アフィン接続の定義式から

$$g(Y,\nabla^*_X Z)=Xg(Y,Z)-g(\nabla_X Y,Z)$$

$$g(Y,\nabla^*_Z X)=Zg(Y,X)-g(\nabla_Z Y,X)$$

であるので，代入して，

$$\begin{aligned}&(\nabla g)(Z,Y,X)+g(Y,T^*(X,Z))\\ &= Zg(Y,X)-g(\nabla_Z Y,X)-g(Y,\nabla_Z X)+Xg(Y,Z)-g(\nabla_X Y,Z)-Zg(Y,X)+g(\nabla_Z Y,X)-g(Y,[X,Z])\\ &= Xg(Y,Z)-g(\nabla_X Y,Z)-g(Y,\nabla_Z X)-g(Y,[X,Z])\end{aligned}$$

であり，したがって，

$$(\nabla g)(X,Y,Z)+g(Y,T(X,Z))=(\nabla g)(Z,Y,X)+g(Y,T^*(X,Z))$$

が成り立つ．したがって，$T=0$, $T^*=0$ が成り立てば,

$$(\nabla g)(X,Y,Z)=(\nabla g)(Z,Y,X)$$

がなりたつ．上記を（あるいは前述の条件 (3) を）コダッチの方程式と呼ぶ.

4.7.3　双対アフィン座標系の存在について

双対平坦な多様体 M では，各点の周りで

$$g\left(\frac{\partial}{\partial x^i},\frac{\partial}{\partial y^j}\right)=\delta_{ij}$$

を満たす局所 $\nabla-$ アフィン座標系 (x^i) と，局所 ∇^*- アフィン座標系 (y^j) を取ることができる．この特性を持った (x^i) と (y^j) を**双対アフィン座標系**と呼ぶ.

証明：　点 $p\in M$ の近傍において ∇ に関するアフィン座標近傍 $(U;\xi^i)$ と ∇^* に関するアフィン座標近傍 $(V;\eta^i)$ とを定める．そして,

$$g_p\left(\left(\frac{\partial}{\partial\xi^i}\right)_p,\left(\frac{\partial}{\partial\eta^j}\right)_p\right)$$

を (i,j) 成分とする行列を $\boldsymbol{G}$ とする．計量の正定値性より，$\boldsymbol{G}$ は正則行列である．ここで，列ベクトル $\boldsymbol{\xi} = [\xi^1,\ldots,\xi^n]^T$，$\boldsymbol{\eta} = [\eta^1,\ldots,\eta^n]^T$ として，

$$\begin{aligned}\boldsymbol{x} &= \boldsymbol{\xi}\\ \boldsymbol{y} &= \boldsymbol{G}\boldsymbol{\eta}\end{aligned}$$

とする新しい座標系を導入する．新しい座標系は $\boldsymbol{x} = [x^1,\ldots,x^n]^T$，$\boldsymbol{y} = [y^1,\ldots,y^n]^T$ であらわした．$\boldsymbol{y} = \boldsymbol{G}\boldsymbol{\eta}$ を成分で表すと，

$$y^j = \sum_k G_{jk}\eta^k \quad \to \quad \eta^k = \sum_j [G^{-1}]_{kj} y^j$$

と表される．したがって，

$$\frac{\partial}{\partial y^j} = \sum_k \frac{\partial \eta^k}{\partial y^j}\frac{\partial}{\partial \eta^k} = \sum_k [G^{-1}]_{kj}\eta^k$$

である．ここで，

$$\begin{aligned}g\left(\left(\frac{\partial}{\partial x^i}\right)_p, \left(\frac{\partial}{\partial y^j}\right)_p\right) g\left(\left(\frac{\partial}{\partial x^i}\right)_p, \left(\sum_k [G^{-1}]_{kj}\frac{\partial}{\partial \eta^k}\right)_p\right)\\ = \sum_k [G^{-1}]_{kj} g\left(\left(\frac{\partial}{\partial x^i}\right)_p, \left(\frac{\partial}{\partial \eta^k}\right)_p\right) = \sum_k [G^{-1}]_{kj} G_{ik} = \delta_{ij}\end{aligned}$$

である．したがって，上で定義した (x^i) と (y^j) が求める座標系である．

以下は必要か？　任意の p で上が証明できていればそれで OK なのではないのか．

こうして作ったアフィン座標系の組 $(x^i), (y^j)$ がすべての $p \in U \cap V$ で

$$g\left(\left(\frac{\partial}{\partial x^i}\right)_p, \left(\frac{\partial}{\partial y^j}\right)_p\right) = \delta_{ij}$$

を満たすことを示す．これを示すには，任意のベクトル場 $X \in \mathfrak{X}(M)$ に対して，

$$Xg\left(\left(\frac{\partial}{\partial x^i}\right)_p, \left(\frac{\partial}{\partial y^j}\right)_p\right) = 0$$

が言えれば，$g\left(\left(\frac{\partial}{\partial x^i}\right)_p, \left(\frac{\partial}{\partial y^j}\right)_p\right) = \delta_{ij}$ がどこでも成り立つことが言える．実際，

$$Xg\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^j}\right) = g\left(\nabla_X \frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^j}\right) + g\left(\frac{\partial}{\partial x^i}, \nabla^*_X \frac{\partial}{\partial y^j}\right)$$

であるが，(x^i) が ∇ アフィン座標系なので

$$\nabla_X \frac{\partial}{\partial x^i} = 0$$

(y^i) が ∇^* アフィン座標系なので

$$\nabla^*_X \frac{\partial}{\partial y^i} = 0$$

である[11]．したがって，$Xg = 0$ であり，$g\left(\left(\frac{\partial}{\partial x^i}\right)_p, \left(\frac{\partial}{\partial y^j}\right)_p\right) = \delta_{ij}$ が $p \in U \cap V$ のどこでも成り立つ．

上の性質を有する局所アフィン座標系 (x^i) と (y^j) の組 $\{(x^i), (y^j)\}$ を双対アフィン座標系と呼ぶ．

[11]補足：

$$\nabla_X \frac{\partial}{\partial x^i} = \nabla_{X^j \frac{\partial}{\partial x^j}} \frac{\partial}{\partial x^i} = X^j {\Gamma^k}_{ji} \frac{\partial}{\partial x^k}$$

個人の感想

１）リーマン幾何学において，2 点を結ぶ最短距離を求めるのは，(i) 与えられた計量と接続を用いて測地線方程式を構成し，これを解いて測地線を求める．(ii) 測地線に沿って計量を積分することにより 2 点間の距離を求める，が一般的な手順である．しかしながら，双対平坦な多様体では，測地線は直線になるが，この直線は最短距離を結ぶ曲線ではない（ようである）．これが本当なら，なぜこんな大事なことを，どの本でも触れてないのか．甘利（[3]）で，ぼそっと結論のみ「つぶやかれて」いるだけである．

したがって，双対平坦な多様体では，距離を計算するのは（少なくとも）大変みたいで，一般的には不可能なのであろう．

２）したがって，ダイバージェンスは双対平坦な多様体に対して，距離の代用品として導入された（らしい）．代用品としては (i) 向きに依存する，(ii) 三角不等式を満たさない，などの問題はあるが，(a) 座標系に依存しない，(b) 2 点が離れていればいるほど大きな値となる（凸関数の性質から証明できる），(c) 2 点を結ぶ経路に依存しない，などの「良い」性質を持っている．

３）ダイバージェンスがわりと良い代用品だとしても以下の疑問はのこる．双対平坦性を放棄して，フィシャー計量とリーマン接続の組み合わせを用いれば，2 点間の距離は計算できる.（どんな場合にも積分が実行できて，距離が簡単に計算できるのかはともかく，少なくとも原理原則は.）これは微分幾何学では標準の方法である．なぜ標準的な方法を用いず，双対平坦性という特殊な考え方を導入するのか（なぜ双対平坦性がそれほど重要なのか,）現時点ではよくわからない.（2 点間の距離を計算する以上に重要な）何か別の理由があるみたいな気もする．

参考文献

主に以下の教科書を参考にした．

[1] 藤原章夫，“情報幾何学の基礎” 牧野書店，2015.

[2] 藤岡敦，“入門情報幾何” 共立出版，2021.

[3] 甘利俊一，“情報幾何学の新展開” サイエンス社，2014.

であるが，アフィン座標系なので ${\Gamma^k}_{ji}$ は恒等的にゼロであり，$\nabla_X \dfrac{\partial}{\partial x^i} = 0$ である．$\nabla^*_X \dfrac{\partial}{\partial y^i} = 0$ についても全く同様な議論が成り立つ．

第 5 章　フィシャー計量とアルファ接続：有限離散確率分布の空間の幾何学を用いた導出

Original:2023-7-03
$(0,3)$ 型テンソルに対する不変性の節より後の節の追加：2023-8-12

有限な離散確率分布の空間における情報幾何学を議論する．確率分布の空間を多様体として考える場合に，計量はある種の変換に対する不変性の要求を満たさなければならない．この “ある種の変換は ”有限な離散確率分布に対するマルコフ埋め込みとして知られている．このノートでは，マルコフ埋め込みから導かれるチャンツォフ (Chentsov) の定理と，この定理から情報幾何にとって基本的なフィシャー計量が導かれることを示す．驚くべきことに，フィシャー計量やアルファ接続などの情報幾何において中心を成す概念は，ここで述べる有限離散確率分布を仮定した議論からでしか導出出来ていないのである．

参考文献 [1],[2] とも，このマルコフ埋め込みからチャンツォフの定理までの部分はわかりにくい. (特に [2] のこの部分の説明はわかりにくく，説明の流れ（論理）が読み取れない.）このノートは比較的ましな [1] の説明に沿って，マルコフ埋め込みとチャンツォフの定理を説明し，フィシャー計量を導く．さらに，有限な離散確率分布の空間における幾何学を議論する．

5.1　対象とする離散確率モデル

まず，基本となる離散確率モデルを定義する．n 個の元からなる有限集合 Ω_n を

$$\Omega_n = \{1, 2, \ldots, n\}$$

として，有限集合の要素である事象（根源事象）が自然数 1 から n でラベル付けされているとする．つまり，事象系を Ω_n で表す．Ω_n 上の確率分布全体の集合を

$$S_{n-1} = \left\{ p(\omega) \,:\, \Omega_n \to \mathbb{R}_{++} \,|\, \sum_{\omega \in \Omega_n} p(\omega) = 1 \right\}$$

と表す．ここで，正の実数の集合を $\mathbb{R}_{++} = \{x \in \mathbb{R} \,|\, x > 0\}$ と表す[1]．ここで，S_{n-1} と表すのは文献 [1] の表し方で，[2] では前節までのように S_n と表している．本ノートでは以降，[1] の表し方 S_{n-1} を用いる．

上記の S_{n-1} の定義は

$$S_{n-1} = \left\{ (p(1), p(2), \ldots, p(n)) \,|\, p(1), p(2), \ldots, p(n) > 0,\, \sum_{i=1}^{n} p(i) = 1 \right\}$$

[1] $\left\{ p \,:\, \Omega_n \to \mathbb{R}_{++} \right\}$ は,「p は Ω_n の元を 1 つ取って正の実数の 1 つに対応づける写像」の意味である．

と書く方がわかりやすいかもしれない．S_{n-1} の各元である確率分布は，n 次元の数ベクトルで表される．したがって，S_{n-1} の各元をベクトル $(p(1), p(2), \ldots, p(n))$ と同一視すれば，集合 S_{n-1} は，$\mathbb{R}^n$ 内の超平面 $\sum_{i=1}^{n} p(i) = 1$ と $\mathbb{R}^n$ の正の象限の共通部分に相当する開領域であるから，自然に $n-1$ 次元多様体とみなせる．

離散確率分布 $p(i) = x^i$, $i \in \Omega_n$ において，各事象に対する確率の値 $(x^1, x^2, \ldots, x^n)$ をこの確率分布のパラメータと考える．したがって，$\boldsymbol{\theta} = (x^1, x^2, \ldots, x^n)$ として，確率分布を $p(i|\boldsymbol{\theta}) = x^i$, $i \in \Omega_n$ と書く場合もある．多様体

$$S_{n-1} = \left\{ p(i|\boldsymbol{\theta}) = x^i,\, i \in \Omega_n \,|\, x^1, \ldots, x^n > 0,\, \sum_{i=1}^{n} x^i = 1 \right\}$$

において，$\boldsymbol{\theta} = (x^1, x^2, \ldots, x^n)$ は，各確率分布を多様体上の点に見立てた場合の点の位置を表す（局所）座標となる．

5.2 マルコフ埋め込み

5.2.1 マルコフ埋め込みの定義

３つの事象にラベルを与えて事象空間を $\Omega_3 = \{1, 2, 3\}$ で表す．3 つの事象 $\{a, b, c\}$ を持つ確率分布の集合 S_{3-1} において，a と b を同一視する，つまり，a あるいは b のどちらかが起こることを事象 $\{z\}$ で表せば，事象空間 $\{z, c\}$ は，2 つの事象を持つ空間となる．

3 つの事象 $\{a, b, c\}$ を持つ確率分布の集合 S_{3-1} の部分集合である事象 $\{z, c\}$ の集合の幾何学構造は，もともと 2 つの事象からなる確率分布の空間の幾何学構造と同じでなければならない．このことを表現するため統計的な同等性を持つ多様体を対応づける写像であるマルコフ埋め込みを以下のように定義する．

定義： n, ℓ を $2 \le n \le \ell$ である自然数とする．以下のように構成される写像

$$\Phi : S_{n-1} \to S_{\ell-1}$$

をマルコフ埋め込みという．ここで，統計モデルは

$$S_{n-1} = \left\{ p(i) = x^i \,(i = 1, \ldots, n)) \,|\, x^1, x^2, \ldots, x^n > 0,\, \sum_{i=1}^{n} x^i = 1 \right\}$$

および，

$$S_{\ell-1} = \left\{ p(i) = y^i \,(i = 1, \ldots, \ell)) \,|\, y^1, y^2, \ldots, y^\ell > 0,\, \sum_{i=1}^{\ell} y^i = 1 \right\}$$

である．

マルコフ埋め込み写像は，数ベクトル $(x^1, \ldots, x^n)$（統計モデル S_{n-1} の元）を入力して，数ベクトル $(y^1, \ldots, y^\ell)$（統計モデル $S_{\ell-1}$ の元）を出力する写像：

$$(y^1, y^2, \ldots, y^\ell) = \Phi(x^1, x^2, \ldots, x^n)$$

である．この写像 Φ は以下のように構成する．

(i) $\Omega_\ell = \{1, 2, \ldots, \ell\}$ を互いに交わらない部分集合の族 $\Omega_\ell = \{C_1, C_2, \ldots, C_n\}$ に分割する．

例： $\Omega_3 = 1, 2, 3$ について $\Omega_3 = \{\{1, 2\}, \{3\}\}$ とする．この場合，

$$C_1 = \{1, 2\}, \quad C_2 = \{3\}$$

である．

(ii) 各 $j : j = 1, \ldots, n$ に対して，C_j に台を持つ Ω_ℓ 上の確率分布を

$$\boldsymbol{Q}_j = (Q_j^1, Q_j^2, \ldots, Q_j^\ell)$$

とする．ここで，$k \in C_j$ なら Q_j^k は正の値を持ち，$k \in \{C_1, \ldots, C_{j-1}, C_{j+1}, \ldots, C_n\}$ なら $Q_j^k = 0$ である．

例： $\ell = 3$，$n = 2$ の例について，

$$\begin{aligned}
&\boldsymbol{Q}_1 = (Q_1^1, Q_1^2, Q_1^3) = (q^1, q^2, 0) \quad \text{ただし} \quad q^1 + q^2 = 1, \quad q^1, q^2 > 0 \\
&\boldsymbol{Q}_2 = (Q_2^1, Q_2^2, Q_2^3) = (0, 0, 1)
\end{aligned}$$

である．

(iii) $(y^1, y^2, \ldots, y^\ell) = \Phi(x^1, x^2, \ldots, y^n)$ を

$$(y^1, y^2, \ldots, y^\ell) = \sum_{j=1}^{n} x^j \boldsymbol{Q}_j = \sum_{j=1}^{n} x^j (Q_j^1, Q_j^2, \ldots, Q_j^\ell)$$

つまり，

$$y^k = \sum_{j=1}^{n} x^j Q_j^k \qquad (k = 1, 2, 3)$$

として計算する．

例： $\ell = 3$，$n = 2$ の例について，計算してみると，

$$y^1 = x^1 Q_1^1 + x^2 Q_2^1 = x^1 q^1 \quad y^2 = x^1 Q_1^2 + x^2 Q_2^2 = x^1 q^2 \quad y^3 = x^1 Q_1^3 + x^2 Q_2^3 = x^2$$

したがって，

$$(y^1, y^2, y^3) = (x^1 q^1, x^1 q^2, x^2) = \Phi(x^1, x^2)$$

である．

この $\ell = 3$，$n = 2$ の例では，3 つの事象 $\Omega_3 = \{1, 2, 3\}$ を，事象 1 か事象 2 が起こることを 1 つの事象とみなした．その結果この「事象 1 か事象 2 が起こる」が起こる確率は $x^1 q^1 + x^1 q^2 = x^1$ となり，この確率モデルは

$$\Phi[S_{2-1}] = \widetilde{S}_{2-1} = \{(q^1 x^1, q^2 x^1, x^2)\} = \{(q^1 x^1, q^2 x^1, x^2) : q^1 + q^2 = 1,\, x^1 + x^2 = 1\}$$

と表される．これは，統計モデル (多様体)

$$S_{3-1} = \{(y^1, y^2, y^3) : y^1,\, y^2,\, y^3 > 0,\, y^1 + y^2 + y^3 = 1\}$$

の特別な場合で，S_{3-1} の部分多様体である．一方，事象の数が 2 個の確率モデルは

$$S_{2-1} = \{(x^1, x^2) : x^1,\, x^2 > 0,\, x^1 + x^2 = 1\}$$

であり，$\widetilde{S}_{2-1}$ と S_{2-1} は同じ統計学的な構造を有してなければならない．したがって，多様体 S_{3-1} の幾何学的構造と多様体 S_{2-1} の幾何学的構造は全く異なるが，多様体 S_{3-1} をこの部分多様体 $\widetilde{S}_{2-1} = \Phi(S_{2-1})$ に制限すれば，$\widetilde{S}_{2-1}$ の幾何学的構造は，S_{2-1} の幾何学的構造に等しくならなければならない．

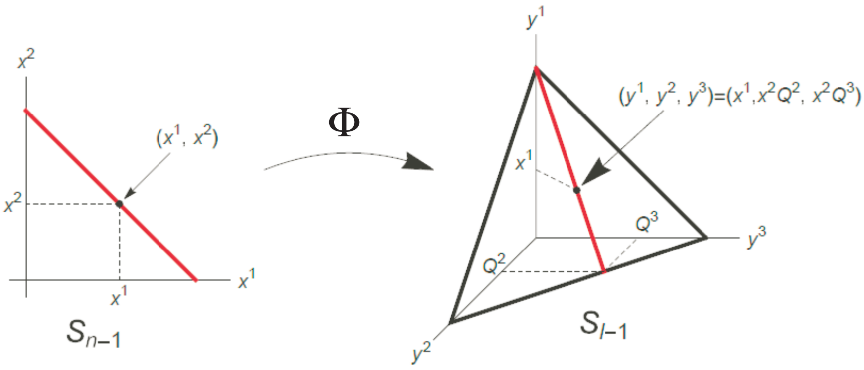

図 5.1: 埋め込まれた部分多様体.

5.2.2　有限・離散確率分布の場合における十分統計量

マルコフ埋め込みと密接に関連した十分統計量を有限・離散な確率分布の場合で定義する.

自然数の部分集合を Ω_ℓ，Ω_n（$\ell \geq n$）として，写像 F を

$$F : \Omega_\ell \to \Omega_n \qquad \Omega_\ell \ni i \mapsto F(i) = j \in \Omega_n \tag{5.1}$$

と定義する．F は全射とする．このとき，

$$q(j|\boldsymbol{\theta}) = \sum_{i=F^{-1}(j)} p(i|\boldsymbol{\theta}) \qquad (i \in \Omega_{¥}\text{え１１，} \quad j \in \Omega_n) \tag{5.2}$$

とおけば，$q(j|\boldsymbol{\theta})$ も確率分布である.（なぜなら，$q(j|\boldsymbol{\theta}) > 0$ と $\sum_{j\in\Omega_n} q(j|\boldsymbol{\theta}) = 1$ が成り立つからである.）

ここで，

$$r(i) = \frac{p(i|\boldsymbol{\theta})}{q(F(i)|\boldsymbol{\theta})} \qquad (i \in \Omega_m)$$

を計算したとき，$r(i)$ が $\boldsymbol{\theta}$ に依存しなければ，$F(i)$ を S に対する十分統計量と呼ぶ.

$r(i)$ が $\boldsymbol{\theta}$ に依存しないことは，確率分布 $p(i)$ と $q(j)$ の $\boldsymbol{\theta}$ 依存性が等しい，つまり，多様体 $\{p(i)\}$ と多様体 $\{q(j)\}$ の幾何構造が等しいことを意味する.

$\ell = 3, n = 2$ の例における具体的な計算例を次の節で示す.

5.2.3　十分統計量の議論との関係

マルコフ埋め込み $\Phi : S_{2-1} \to S_{3-1}$ は，S_{2-1} を S_{3-1} に埋め込み，S_{3-1} の部分多様体 $\Phi[S_{2-1}] = \widetilde{S}_{2-1}$ を作り出す写像である.

S_{3-1} の部分多様体 $\widetilde{S}_{2-1}$ と，S_{2-1} は統計学的には同等であり，多様体として同じ幾何学的構造を持っている．このことは，統計学では「$\widetilde{S}_{2-1}$ は十分統計量を持つ」と表す.

それでは，これら S_{3-1}，$\widetilde{S}_{2-1}$，S_{2-1} は，十分統計量の議論のどこに対応するのであろうか？

われわれが観測するのは S_{3-1} からのサンプル y^1, y^2, y^3 であるが，確率データ y^1, y^2, y^3 が S_{3-1} の全体からではなく，部分多様体 $\widetilde{S}_{2-1}$ からのサンプルであったとするシナリオを考える．

$$(p(1), p(2), p(3)) = (y^1, y^2, y^3) = (q^1x^1, q^2x^1, x^2)$$

が観測されたとする．ここで，写像 $F : \Omega_3 \to \Omega_2$ を

$$F(1) = 1$$
$$F(2) = 1$$
$$F(3) = 2$$

として，この F を用いれば，式 (5.2) の確率分布 $q(j) : j \in \Omega_2$ を

$$q(1) = \sum_{i \in F^{-1}(1)} p(i) = q^1x^1 + q^2x^1 = x^1$$
$$q(2) = \sum_{i \in F^{-1}(2)} p(i) = x^2$$

として計算できる．したがって，F が十分統計量であるかどうかの評価のため，$\gamma(i)$ を計算してみると，

$$\gamma(1) = \frac{p(1)}{q(F(1))} = \frac{q^1x^1}{x^1} = q^1 \quad \gamma(2) = \frac{p(2)}{q(F(2))} = \frac{q^2x^1}{x^1} = q^2 \quad \gamma(3) = 1$$

であり，$\gamma(i)$ はパラメータ $\{(x^1, x^2)\}$ には依存せず，$F(i)$ は十分統計量である．

ここで，F を確率分布に作用する写像とみなして，

$$F[\Phi[S_{2-1}]] = F[\widetilde{S}_{2-1}] = F[\{(p(1), p(2), p(3))\}] = \{(q(1), q(2))\} = \{(x^1, x^2)\}$$

であるので，結局，

$$F[\Phi[S_{2-1}]] = S_{2-1}$$

である．つまり，十分統計量 F を見つけることは，$\widetilde{S}_2$ からの確率データをもとにして，高次元多様体 S_{3-1} に埋め込まれている低次元多様体 S_{2-1} を見つけることに等しい．十分統計量 F は埋め込み写像の逆像になっている．すなわち，

$$F \circ \Phi = I \quad (= \text{恒等写像}) \tag{5.3}$$

である．反対に言えば，Φ がマルコフ埋め込みであれば，式 (5.3) を満たす F は必ず存在する，つまり，F は必ず十分統計量である．

5.3　マルコフ埋め込みの例

第 5.4 節でチャンツォフの定理の証明に用いる 3 種類のマルコフ埋め込みの例を説明する．

5.3.1　マルコフ埋め込み：$\mathbb{R}^n_{++} \to \mathbb{R}^\ell_{++}$

先に述べたように，離散統計モデルは

$$S_{n-1} = \left\{(x^1, x^2, \ldots, x^n) \,|\, x^1, x^2, \ldots, x^n > 0,\, \sum_{i=1}^n x^i = 1\right\}$$

と定義されるが，写像の微分の計算においては，$\sum_{i=1}^n x^i = 1$ の制限を外したモデル，

$$\left\{(x^1, x^2, \ldots, x^n) \,|\, x^1, x^2, \ldots, x^n > 0\right\}$$

に対するマルコフ埋め込みを考える．上記は $\mathbb{R}^n_{++}$ に等しいので，

$$\mathbb{R}^n_{++} = \left\{(x^1, x^2, \ldots, x^n) \,|\, x^1, x^2, \ldots, x^n > 0\right\}$$

と表す．マルコフ埋め込み

$$\Phi : S_{n-1} \to S_{\ell-1} \qquad (n \le \ell)$$

を

$$\Phi : \mathbb{R}^n_{++} \to \mathbb{R}^\ell_{++} \qquad (n \le \ell)$$

と拡張して考える．

5.3.2　例 1：並べ替え

写像と十分統計量

以下のマルコフ埋め込みを考える．

$$\Phi : \Omega_n \to \Omega_n$$

で全単射とする．つまり，$\Omega_n = \{1, 2, \ldots, n\}$ に対して，Φ はこれらを並べ替えたどれかに移す．すなわち，番号を並べ替えた確率分布をつくる．

確率変数の番号には意味が無いので，並べ替えても確率分布は変化しない．これを示すために十分統計量の確率分布を求めてみる．

この例の場合，$n = \ell$ であるので，

$$S_{n-1} = \left\{(x^1, x^2, \ldots, x^n) \,|\, x^1, x^2, \ldots, x^n > 0,\, \sum_{i=1}^n x^i = 1\right\}$$

および，

$$\Phi(S_{n-1}) = \left\{(y^1, y^2, \ldots, y^n) \,|\, y^1, y^2, \ldots, y^n > 0,\, \sum_{i=1}^\ell y^i = 1\right\}$$

であり，

$$(y^1, y^2, \ldots, y^n) = \Phi(x^1, x^2, \ldots, x^n)$$

であるが，$y^1, y^2, \ldots, y^n$ は $x^1, x^2, \ldots, x^n$ を並べ変えた結果である．$1, 2, \ldots, n$ の並べ替え結果を $\zeta(1), \zeta(2), \ldots, \zeta(n)$ と表すと[2]，

$$(x^{\zeta(1)}, x^{\zeta(2)}, \ldots, x^{\zeta(n)}) = \Phi(x^1, x^2, \ldots, x^n)$$

である．

ここで，

$$\Phi(S_{n-1}) = \{(\widetilde{p}(1), \widetilde{p}(2), \ldots, \widetilde{p}(n)) \,|\, \widetilde{p}(i) = x^{\zeta(i)}\}$$

と書くことにする．さらに，写像 F により，統計モデル $\Phi(S_{n-1})$ が $F[\Phi(S_{n-1})]$ に変換されるとすれば，

$$F[\Phi(S_{n-1})] = \{q(1), \ldots, q(n) \,|\, (q(1), \ldots, q(n)) = F(\widetilde{p}(1), \ldots, \widetilde{p}(n))\}$$

である．ここで，F として Φ^{-1} を用いれば，

$$\begin{aligned}(q(1), \ldots, q(n)) = F(\widetilde{p}(1), \ldots, \widetilde{p}(n)) = \Phi^{-1}(\widetilde{p}(1), \ldots, \widetilde{p}(n)) &= \Phi^{-1}(x^{\zeta(1)}, x^{\zeta(2)}, \ldots, x^{\zeta(n)}) \\ &= (x^{\zeta^{-1}(\zeta(1))}, x^{\zeta^{-1}(\zeta(2))}, \ldots, x^{\zeta^{-1}(\zeta(n))}) = (x^1, x^2, \ldots, x^n) \quad (5.4)\end{aligned}$$

である．ここで，$\zeta^{-1}(i)$ は，i に対する並べ替え $\zeta(i)$ の逆写像を意味する．並べ替えで添え字が $i \to \zeta(i)$ と変わり，逆の並べ替えで $\zeta^{-1}(\zeta(i)) = i$ となる．したがって，十分統計量の評価関数は，

$$\frac{\widetilde{p}(i)}{q(F(i))} = \frac{\widetilde{p}(i)}{p(F^{-1}(F(i)))} = \frac{\widetilde{p}(i)}{p(\Phi(\Phi^{-1}(i)))} = \frac{\widetilde{p}(i)}{p(i)} = \frac{x^{\zeta(i)}}{x^i}$$

と表され，上記評価関数はパラメータ x^i に依存してしまい $\widetilde{p}(i)$ は十分統計量ではなが，$x^1 = x^2 = \cdots = x^n$ の場合のみ上記評価関数がパラメータに依存せず十分統計量となる．

写像の微分の計算

マルコフ埋め込み

$$(y^1, y^2, \ldots, y^\ell) = \Phi(x^1, x^2, \ldots, x^n)$$

に対して，写像の微分 $d\Phi$ はヤコビ行列

$$[d\Phi]_{ij} = \frac{\partial y^i}{\partial x^j}$$

を計算するのであるが，やっかいなのは，Φ が確率モデル間の写像

$$\Phi : S_n \to S_\ell$$

であるとすれば，制約条件 $\sum_{i=1}^{\ell} y^i = 1$ を考慮してヤコビ行列を計算しなければならないことである．そこで．本節においては，ひとまず，マルコフ埋め込みが

$$\Phi : \mathbb{R}^n_{++} \to \mathbb{R}^\ell_{++}$$

と，拡張されているとしてヤコビ行列を計算する．(確率モデルとしての制約は，後ほどチェンツォフの定理の証明において議論する．)

[2]左から 1 番目に来る番号を $\zeta(1)$ として，2 番目に来る番号を $\zeta(2)$ として，... とする．$\{1, 2, 3\} \to \{2, 3, 1\}$ であれば，$\zeta(1) = 2$，$\zeta(2) = 3$，$\zeta(3) = 1$ である．ちなみに，逆写像は $\zeta^{-1}(2) = 1$，$\zeta^{-1}(3) = 2$，$\zeta^{-1}(1) = 3$ である．

すると，並べ替えに対する埋め込み写像 Φ は,

$$\Phi(x^1, x^2, \ldots, x^n) = (x^{\zeta(1)}, x^{\zeta(2)}, \ldots, x^{\zeta(n)}) = (y^1, \ldots, y^n)$$

で表されるので，写像の微分 $d\Phi$ を表すヤコビ行列は

$$d\Phi = \begin{bmatrix} \dfrac{\partial x^{\zeta(1)}}{\partial x^1} & \dfrac{\partial x^{\zeta(2)}}{\partial x^1} & \cdots & \dfrac{\partial x^{\zeta(n)}}{\partial x^1} \\ \dfrac{\partial x^{\zeta(1)}}{\partial x^2} & \dfrac{\partial x^{\zeta(2)}}{\partial x^2} & \cdots & \dfrac{\partial x^{\zeta(n)}}{\partial x^2} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial x^{\zeta(1)}}{\partial x^n} & \dfrac{\partial x^{\zeta(2)}}{\partial x^n} & \cdots & \dfrac{\partial x^{\zeta(n)}}{\partial x^n} \end{bmatrix} \tag{5.5}$$

である．

簡単な例で考えてみると，$\Omega = \{1, 2, 3\}$ として,

$$\Phi : \{1, 2, 3\} \to \{2, 3, 1\}$$

の例で考える．

$$\Phi : \Omega_3 \to \Omega_3 : \Phi(x^1, x^2, x^3) = (x^2, x^3, x^1)$$

であるので，この写像 Φ の微分 $d\Phi$ はヤコビ行列

$$d\Phi = \begin{bmatrix} \dfrac{\partial x^2}{\partial x^1} & \dfrac{\partial x^3}{\partial x^1} & \dfrac{\partial x^1}{\partial x^1} \\ \dfrac{\partial x^2}{\partial x^2} & \dfrac{\partial x^3}{\partial x^2} & \dfrac{\partial x^1}{\partial x^2} \\ \dfrac{\partial x^2}{\partial x^3} & \dfrac{\partial x^3}{\partial x^3} & \dfrac{\partial x^1}{\partial x^3} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{5.6}$$

である．このヤコビ行列のランクは，$d\Phi$ が全単射なので 3 になる．全く同じ考え方で，式 (5.5) のヤコビ行列はランク n となる．したがって，$d\Phi$ は全単射であり，Φ は埋め込みである．

ℓ 番目要素のみ 1 で他の要素はゼロである 1×3 の行ベクトルを $\widetilde{\boldsymbol{e}}_\ell$ として用いると，式 (5.6) に示す写像の微分 $d\Phi$ は,

$$d\Phi = \begin{bmatrix} \widetilde{\boldsymbol{e}}_{\zeta^{-1}(1)} \\ \widetilde{\boldsymbol{e}}_{\zeta^{-1}(2)} \\ \widetilde{\boldsymbol{e}}_{\zeta^{-1}(3)} \end{bmatrix}$$

と表すことができる.（逆写像が $\zeta^{-1}(2) = 1$，$\zeta^{-1}(3) = 2$，$\zeta^{-1}(1) = 3$ であるので，簡単に確認できる.）一般の場合，S_{n-1} における写像の微分 $d\Phi$ は，$\widetilde{\boldsymbol{e}}_\ell$ を ℓ 番目要素のみ 1 で他の要素はゼロである $1 \times n$ の行ベクトルを表すとして,

$$d\Phi = \begin{bmatrix} \widetilde{\boldsymbol{e}}_{\zeta^{-1}(1)} \\ \vdots \\ \widetilde{\boldsymbol{e}}_{\zeta^{-1}(n)} \end{bmatrix} \tag{5.7}$$

と表すことができる．

5.3.3 例 2：補間による事象空間の拡大

写像と十分統計量

$$S_{n-1} = \left\{ p(i) = x^i \,|\, p(i)\,(i = 1, \ldots, n) > 0,\, \sum_{i=1}^{n} p(i) = 1 \right\}$$

に対し，ある自然数 m として，事象空間を m 倍に拡大する

$$\Phi(x^1, \ldots, x^n) = \Big(\underbrace{\frac{x^1}{m}, \ldots, \frac{x^1}{m}}_{m\text{ 個}}, \underbrace{\frac{x^2}{m}, \ldots, \frac{x^2}{m}}_{m\text{ 個}}, \ldots, \underbrace{\frac{x^n}{m}, \ldots, \frac{x^n}{m}}_{m\text{ 個}}\Big) = (y^1, y^2, \ldots, y^{mn})$$

という，マルコフ埋め込みを考える．すなわち，埋め込まれた部分多様体は

$$\Phi(S_{n-1}) = \left\{ \widetilde{p}(i) = y^i \,|\, \widetilde{p}(i)\,(i = 1, \ldots, nm) > 0,\, \sum_{i=1}^{nm} \widetilde{p}(i) = 1 \right\}$$

と表すことができる．

ここで，写像 $F : \Omega_{nm} \to \Omega_n$ を

$$F(i) = j \quad \text{ただし，} i \in [(j-1)m+1, (j-1)m+2, \ldots, jm]\text{，また } j = 1, \ldots, n \text{ である．}$$

つまり，

$$\begin{aligned}
&F(1) = 1, \cdots, F(m) = 1 \\
&F(m+1) = 2, \cdots, F(2m) = 2 \\
&F(2m+1) = 3, \cdots, F(3m) = 3 \\
&\qquad\quad \vdots \qquad\qquad\qquad\quad \vdots \\
&F((n-1)m+1) = n, \cdots, F(nm) = n
\end{aligned}$$

と定義する．F は全射である．すると，$j = 1, 2, \ldots, n$ に対して，式 (5.2) に示す $q(j)$ は，

$$q(j) = \sum_{i \in F^{-1}(j)} \widetilde{p}(i) = \sum_{i=(j-1)m+1}^{jm} \widetilde{p}(i) = x^j$$

として求まる．この確率分布を元に持つ統計モデル $F[\Phi(S_{n-1})]$ は，

$$F[\Phi(S_{n-1})] = \left\{ q(1), \ldots, q(n) \,|\, q(j) = \sum_{i=(j-1)m+1}^{jm} \widetilde{p}(i) = x^j \right\}$$

となる．すなわち，

$$F[\Phi(S_{n-1})] = S_{n-1}$$

であり，十分統計量の評価関数は，$i \in [(j-1)m+1, jm]$ であれば，$\widetilde{p}(i) = x^j/m$ でありるので

$$r(i) = \frac{\widetilde{p}(i)}{q(F(i))} = \frac{\widetilde{p}(i)}{q(j)} = \frac{x^j/m}{x^j} = \frac{1}{m}$$

を得る．したがって，$F(\cdot)$ は十分統計量である．

写像の微分

この例におけるマルコフ埋め込み写像は

$$\Phi(x^1,\dots,x^n)=\Big(\underbrace{\frac{x_1}{m},\dots,\frac{x_1}{m}}_{m\text{ 個}},\underbrace{\frac{x_2}{m},\dots,\frac{x_2}{m}}_{m\text{ 個}},\dots,\underbrace{\frac{x_n}{m},\dots,\frac{x_n}{m}}_{m\text{ 個}}\Big)=(y^1,y^2,\dots,y^{mn})$$

である．マルコフ埋め込み Φ が

$$\Phi:\mathbb{R}^n_{++}\to\mathbb{R}^\ell_{++}\qquad(n\le\ell)$$

と定義されているとする．すると，この写像の微分（すなわちヤコビ行列）は，$n\times(nm)$ の行列で

$$d\Phi=\begin{bmatrix}\dfrac{\partial y^1}{\partial x^1} & \dfrac{\partial y^2}{\partial x^1} & \cdots & \dfrac{\partial y^{nm}}{\partial x^1}\\ \dfrac{\partial y^1}{\partial x^2} & \dfrac{\partial y^2}{\partial x^2} & \cdots & \dfrac{\partial y^{nm}}{\partial x^2}\\ \vdots & \vdots & \ddots & \vdots\\ \dfrac{\partial y^1}{\partial x^n} & \dfrac{\partial y^2}{\partial x^n} & \cdots & \dfrac{\partial y^{nm}}{\partial x^n}\end{bmatrix}\tag{5.8}$$

となる．ここで，要素が全て 1 である $1\times m$ の行ベクトルを

$$\mathbf{1}_m=(\underbrace{1,1,\dots,1}_{m\text{ 個}})$$

と表し，要素が全て 0 である $1\times m$ の行ベクトルを

$$\mathbf{0}_m=(\underbrace{0,0,\dots,0}_{m\text{ 個}})$$

と表せば，

$$d\Phi=\begin{bmatrix}\frac{1}{m}\mathbf{1}_m & \mathbf{0}_m & \cdots & \mathbf{0}_m\\ \mathbf{0}_m & \frac{1}{m}\mathbf{1}_m & \cdots & \mathbf{0}_m\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0}_m & \mathbf{0}_m & \cdots & \frac{1}{m}\mathbf{1}_m\end{bmatrix}\tag{5.9}$$

と表すことができる．$d\Phi$ は $n\times(nm)$ の行列であり，各行は線形独立であるので，$d\Phi$ のランクは n である．したがって，$d\Phi$ は単射であり，Φ ははめ込みである．

$d\Phi$ の表し方として，式 (5.9) 以外を考える．$1\times(nm)$ である行ベクトルで，ℓ 番目の要素のみが 1 で他の要素はすべてゼロであるものを

$$\widetilde{\boldsymbol{e}}_\ell=(0,0,\dots,0,\underbrace{1}_{\ell\text{番目}},0,\dots,0)$$

と書くことにする．一方，式 (5.9) の第 k 行を行ベクトルとして書くと，

$$\text{式 (5.9) の第 }k\text{ 行}=\frac{1}{m}(0,0,\dots,0,\underbrace{1,\dots,1}_{k(m-1)+1\sim km\text{ 番目}},0,\dots,0)$$

となる．すなわち，この第 k 行は，第 $k(m-1)+1$ 番目 から第 km 番目 までの要素が 1 で他の要素が全てゼロである行ベクトルである．この第 k 行は，$\widetilde{\boldsymbol{e}}_\ell$ を用いて表すと，

$$\text{式 (5.9) の第}\,k\,\text{行} = \frac{1}{m}\sum_{\ell=(k-1)m+1}^{km}\widetilde{\boldsymbol{e}}_\ell$$

とあらわされるので，結局，$d\Phi$ は

$$d\Phi = \begin{bmatrix} \frac{1}{m}\sum_{\ell=1}^{m}\widetilde{\boldsymbol{e}}_\ell \\ \frac{1}{m}\sum_{\ell=m+1}^{2m}\widetilde{\boldsymbol{e}}_\ell \\ \vdots \\ \frac{1}{m}\sum_{\ell=(k-1)m+1}^{nm}\widetilde{\boldsymbol{e}}_\ell \end{bmatrix} \tag{5.10}$$

と表すことができる．この表現はチャンツォフ定理の証明で用いる．

5.3.4 例 3：補間による事象空間の拡大 (II)

写像と十分統計量

前節のマルコフ埋め込み写像を少し一般化したもの．

$$\Omega_n = \{1, 2, \ldots, n\}$$

に対して，ℓ 個の要素を持った Ω_ℓ を考える．ここで，

$$\ell = m_1 + m_2 + \cdots + m_n$$

として，Ω_ℓ を

$$\Omega_\ell = \{0, 1, 2, \ldots, m_1,\ m_1+1, \ldots, (m_1+m_2),\ (m_1+m_2)+1, \ldots, \sum_{j=1}^{n-1} m_j,\ \sum_{j=1}^{n-1} m_j + 1, \ldots, \ell\}$$

とする．ここで，

$$\Phi(x^1, \ldots, x^n) = \Big(\underbrace{\frac{x^1}{m_1}, \ldots, \frac{x^1}{m_1}}_{m_1\text{個}}, \underbrace{\frac{x^2}{m_2}, \ldots, \frac{x^2}{m_2}}_{m_2\text{個}}, \ldots, \underbrace{\frac{x^n}{m_n}, \ldots, \frac{x^n}{m_n}}_{m_n\text{個}}\Big) = (y^1, y^2, \ldots, y^\ell)$$

という，マルコフ埋め込みを考える．

すなわち，埋め込まれた部分多様体は，上式の y^i を用いて

$$\Phi(S_{n-1}) = \left\{\widetilde{p}(i) = y^i \,|\, \widetilde{p}(i)\,(i = 1, \ldots, \ell) > 0,\ \sum_{i=1}^{\ell}\widetilde{p}(i) = 1\right\}$$

である．ここで，写像 $F:\Omega_\ell\to\Omega_n$ を

$$
\begin{aligned}
&F(0)=0\\
&F(1)=1,\cdots,F(m_1)=1\\
&F(m_1+1)=2,\cdots,F(m_1+m_2)=2\\
&F(m_1+m_2+1)=3,\cdots,F(m_1+m_2+m_3)=3\\
&\qquad\quad\vdots\qquad\qquad\qquad\qquad\vdots\\
&F(\sum_{j=1}^{n-1}m_j+1)=n,\cdots,F(\ell)=n
\end{aligned}
$$

と定義する．F は全射である．すると，$j=1,2,\dots,n$ に対して，

$$
q(j)=\sum_{i\in F^{-1}(j)}\widetilde{p}(i)=\sum_{i=\sum_{j=1}^{j-1}m_j+1}^{\sum_{j=1}^{j}m_j}\frac{x^j}{m_j}=m_j\frac{x^j}{m_j}=x^j
$$

である．$q(j)$ が確率分布であることは（簡単に）示すことができる．この確率分布を元に持つ統計モデル（多様体）を $F[\Phi(S_{n-1})]$ と書くことにすれば，

$$
F[\Phi(S_{n-1})]=\left\{q(1),\dots,q(n)\,|\,q(j)=\sum_{i=\sum_{j=1}^{j-1}m_j+1}^{\sum_{j=1}^{j}m_j}\widetilde{p}(i)=x^j\right\}
$$

となる．すなわち，

$$
F[\Phi(S_{n-1})]=S_{n-1}
$$

であり，十分統計量の評価関数は $i\in[\sum_{j=1}^{j-1}m_j+1,\sum_{j=1}^{j}m_j]$ の場合，

$$
r(i)=\frac{\widetilde{p}(i)}{q(F(i))}=\frac{\widetilde{p}(i)}{q(j)}=\frac{x_j/m_j}{x_j}=\frac{1}{m_j}
$$

となるので，$F(\cdot)$ は十分統計量である．

写像の微分

マルコフ埋め込み Φ が

$$
\Phi:\mathbb{R}^n_{++}\to\mathbb{R}^\ell_{++}\qquad(n\le\ell)
$$

と定義されているとすれば，この写像の微分（ヤコビ行列）は，$n\times\ell$ の行列で

$$
d\Phi=\begin{bmatrix}
\frac{\partial y^1}{\partial x^1} & \frac{\partial y^2}{\partial x^1} & \cdots & \frac{\partial y^\ell}{\partial x^1}\\
\frac{\partial y^1}{\partial x^2} & \frac{\partial y^2}{\partial x^2} & \cdots & \frac{\partial y^\ell}{\partial x^2}\\
\vdots & \vdots & \ddots & \vdots\\
\frac{\partial y^1}{\partial x^n} & \frac{\partial y^2}{\partial x^n} & \cdots & \frac{\partial y^\ell}{\partial x^n}
\end{bmatrix}
=\begin{bmatrix}
\frac{1}{m_1}\mathbf{1}_{m_1} & \mathbf{0}_{m_2} & \cdots & \mathbf{0}_{m_n}\\
\mathbf{0}_{m_1} & \frac{1}{m_2}\mathbf{1}_{m_2} & \cdots & \mathbf{0}_{m_n}\\
\vdots & \vdots & \ddots & \vdots\\
\mathbf{0}_{m_1} & \mathbf{0}_{m_2} & \cdots & \frac{1}{m_n}\mathbf{1}_{m_n}
\end{bmatrix}
\tag{5.11}
$$

と表すことができる．$d\Phi$ は $n\times\ell$ の行列（$n<\ell$）であり，各行は線形独立であるので，$d\Phi$ のランクは n である．したがって，$d\Phi$ は単射であり，Φ ははめ込みである．

行ベクトル $\widetilde{\boldsymbol{e}}_\ell$ を用いて表すと

$$d\Phi=\begin{bmatrix}\frac{1}{m_1}\sum_{\ell=1}^{m_1}\widetilde{\boldsymbol{e}}_\ell\\ \frac{1}{m_2}\sum_{\ell=m_1+1}^{m_1+m_2}\widetilde{\boldsymbol{e}}_\ell\\ \vdots\\ \frac{1}{m_n}\sum_{\ell=(m_1+\cdots+m_{n-1}+1)}^{m_1+\cdots+m_n}\widetilde{\boldsymbol{e}}_\ell\end{bmatrix} \tag{5.12}$$

と表すことができる．この表現はチャンツォフ定理の証明で用いる．

5.3.5 引き戻しと誘導計量

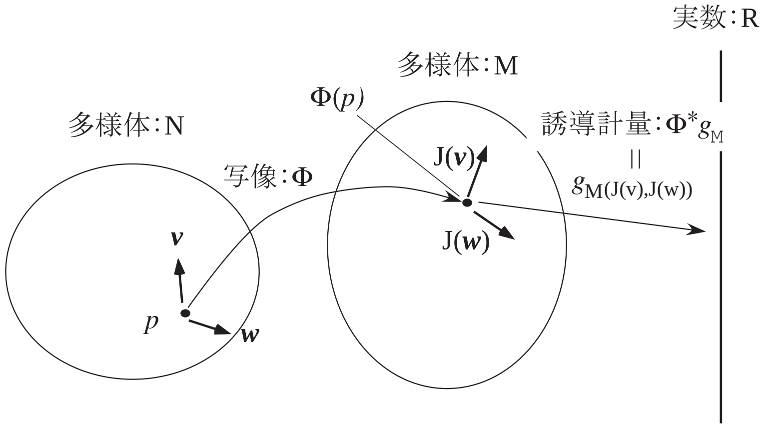

図 5.2: 誘導計量の概念.

M，N を多様体とする．M には計量 g_M が与えられているとして，$\Phi: N\to M$ をはめ込み（埋め込み）とする．ここで，$N\le M$ である．Φ を用いて N に計量を与えることができる．すなわち，$p\in N$，$\boldsymbol{v},\boldsymbol{w}\in T_pN$ に対して，Φ と g_M を用いて，以下のように N に計量を定義することができる．この計量を (Φ^*g_M) と書くと，

$$(\Phi^*g_M)_p(\boldsymbol{v},\boldsymbol{w})=(g_M)_{\Phi(p)}((d\Phi)_p(\boldsymbol{v}),(d\Phi)_p(\boldsymbol{w}))=(g_M)_{\Phi(p)}(J_p(\Phi)(\boldsymbol{v}),J_p(\Phi)(\boldsymbol{w}))$$

である．ここで，$(d\Phi)_p$ は，写像 Φ の微分で，$J_p(\Phi)$ は Φ の微分に対応したヤコビ行列である．$(\Phi^*g_M)_p$ を Φ による誘導計量，あるいは，Φ による g_M の引き戻しと呼ぶ．(図 5.2 参照．)

5.3.6 マルコフ埋め込み：まとめ

有限離散な確率分布を前提とする．

- N 次元と M 次元（$N\le M$）統計モデルを

  $S_N=\{p(i)=x^i\,|\,\boldsymbol{x}=(x^1,\ldots,x^N)\in\Theta_N,\,i\in\Omega_N\}\qquad S_M=\{p(i)=y^i\,|\,\boldsymbol{y}=(y^1,\ldots,y^M)\in\Theta_M\,i\in\Omega_M\}$

と定める．ここで，$\boldsymbol{x}$ を N 次元パラメータ，Θ_N を N 次元パラメータ $\boldsymbol{x}$ の空間，$\boldsymbol{y}$ は M 次元パラメータ，Θ_M は M 次元パラメータ $\boldsymbol{y}$ の空間である[3]．

- つまり，S_N と S_M は，それぞれ，事象空間 $\Omega_N = \{1, 2, \ldots, N\}$ と $\Omega_M = \{1, 2, \ldots, M\}$ 上の統計モデルであるとする．
- S_N と Θ_N，S_M と Θ_M は，対応する元を考えることにより同一視できる．(Θ_N における点 $\boldsymbol{x}$ は，S_N における $\boldsymbol{x}$ をパラメータに持つ確率分布 $p(i|\boldsymbol{x})$ が表す点と同一視できるという意味である．）
- 写像 $\Phi : \mathbb{R}^N \to \mathbb{R}^M$，つまり，$N$ 次元数ベクトルに作用し，M 次元数ベクトルを出力する写像：

$$(y^1, y^2, \ldots, y^M) = \Phi(x^1, x^2, \ldots, x^N)$$

を用いれば，$\Phi : \Theta_N \ni \boldsymbol{x} \mapsto \Phi(\boldsymbol{x}) \in \Theta_M$ であり，

$$\Phi(\Theta_N) = \{\Phi(\boldsymbol{x}) \,|\, \boldsymbol{x} \in \Theta_N\}$$

と書けば

$$\Phi(\Theta_N) \subset \Theta_M$$

となる．(図 5.3(a) を参照のこと．）つまり，$\Phi(\Theta_N)$ は Θ_M の部分空間を成している．

- ここで，$\boldsymbol{x} \in \Theta_N$ と $\Phi(\boldsymbol{x}) \in \Theta_M$ を，S_N と S_M の元 $p(i|\boldsymbol{x})$ と $p(i|\Phi(\boldsymbol{x}))$ に対応させることにより，Φ は S_N から S_M への写像と考えることもできる．ここで，$\Phi(\Theta_N)$ に対応した Ω_M 上の N 次元統計モデル $\Phi(S_N)$ を

$$\Phi(S_N) = \{p(i|\Phi(\boldsymbol{x}))|i \in \Omega_M, \Phi(\boldsymbol{x}) \in \Theta_M\}$$

と定める．図 5.3(b) に示すように，$\Phi(\Theta_N) \subset \Theta_M$ に対応して $\Phi(S_N) \subset S_M$ である．この $\Phi(S_N)$ が，写像 Φ によって S_M に埋め込まれた部分多様体である．(後で証明するが，Φ がマルコフ埋め込み写像なら，$\Phi(S_N)$ の幾何学的構造は，S_N の幾何学的構造に等しい．）

- 「$\Phi(S_N)$ の幾何学的構造は，S_N の幾何学的構造に等しい」と言うことは，統計学的には「$\Phi(S_N)$ は十分統計量を持つ」事と等価である．

これについて説明するため，$\Phi(S_N)$ に関する十分統計量を求める．S_N と S_M に対応する事象空間 Ω_N と Ω_M の間に全射

$$F : \Omega_M \to \Omega_N \quad | \quad \Omega_M \ni i \mapsto j = F(i) \in \Omega_N$$

を定める．(F は Φ とは反対向きの写像である．$M \geq N$ に注意．図 5.3(c) を参照のこと．）すると，

$$q(j|\Phi(\boldsymbol{x})) = \sum_{i \in F^{-1}(j)} p(i|\Phi(\boldsymbol{x})) \qquad (j \in \Omega_N, \quad i \in \Omega_M)$$

は Ω_N 上の確率分布である．この $q(j|\Phi(\boldsymbol{x}))$ は，Ω_N 上の統計モデルを形成し，

$$F(\Phi(S_N)) = \{q(j|\Phi(\boldsymbol{x}))|j \in \Omega_N, \boldsymbol{\theta} \in \Theta_N\}$$

[3] 前節までの書き方では，

$$S_{N-1} = \{p(i) = x^i \,|\, \boldsymbol{x} = (x^1, \ldots, x^N) \in \Theta_N, x^1, \ldots, x^N \geq 0, \sum_{i=1}^{N} x^i = 1\}$$

と表していたものであるが，簡単さのためこのようにした．

と書くことにすれば，Φ **がマルコフ埋め込みで** F **がその逆像となっていれば，**$q(j|\Phi(\boldsymbol{x}))$ からなる統計モデル（多様体）は，Ω_M に "埋め込まれていた" N 次元統計モデル $\Phi(S_N)$ を，写像 F で Ω_N 上に戻したものである．つまり，$F \circ \Phi = I$ が成り立ち

$$F(\Phi(S_N)) = S_N$$

となる．

ここで，

$$r(i) = \frac{p(i|\Phi(\boldsymbol{x}))}{q(F(i)|\Phi(\boldsymbol{x}))} \qquad (\boldsymbol{x} \in \Theta_N, i \in \Omega_M) \tag{5.13}$$

として計算した $r(i)$ が $\boldsymbol{x}$ に依存しない時，F は十分統計量であると言われる．F が十分統計量であるとき，写像 Φ をマルコフうめ込みという．

式 (5.13) の分子と分母に関して，$\Phi(S_N) = \{p(i|\Phi(\boldsymbol{x}))\}$（$S_M$ 内の部分多様体）であり，$\{q(F(i)|\Phi(\boldsymbol{x}))\} = F(\Phi(S_N)) = S_N$ であるので，$r(i)$ がパラメータに依存しない（F が十分統計量である）ことは，$\Phi(S_N)$（S_M 内の部分多様体）の座標 $\boldsymbol{x}$ への依存の仕方（多様体としての幾何構造）と，S_N の座標 $\boldsymbol{x}$ への依存の仕方（多様体としての幾何構造）が同じ，つまり，

$$S_N = \Phi(S_N)$$

が成り立つ事を意味する．(図 5.3(d) を参照.)

もう 1 度議論をまとめると，以下のようになる．

(A) まず，

$$S_N = \{p(j) = x^j \,|\, \boldsymbol{x} \in \Theta_N,\, j \in \Omega_N\}$$

と S_N を定義する．

(B) 第 5.2.1 節に述べたようにマルコフ埋め込み Φ を定義する．

(C) $\Phi(S_N) = \{p(i|\Phi(\boldsymbol{x}))|i \in \Omega_M, \boldsymbol{x} \in \Theta_N\}$ を計算すれば，これは，S_N と同じ多様体としての幾何構造（座標 $\boldsymbol{x}$ への依存の仕方）を持っている．このことは次のように示される．

(D) Φ がマルコフ埋め込みであれば，必ず逆像 F が存在し，Ω_N 上の確率分布

$$q(j|\Phi(\boldsymbol{x})) = \sum_{i \in F^{-1}(j)} p(i|\Phi(\boldsymbol{x})) \qquad (j \in \Omega_N, \quad i \in \Omega_M)$$

を求めることができる．この $q(j|\Phi(\boldsymbol{x}))$ は，Ω_N 上の統計モデル

$$F(\Phi(S_N)) = \{q(j|\Phi(\boldsymbol{x}))|j \in \Omega_N, \boldsymbol{\theta} \in \Theta_N\}$$

を構成し，$F \circ \Phi = I$（I は恒等写像を意味する）であるので

$$S_N = F(\Phi(S_N))$$

が成り立つ．つまり，$F(\Phi(S_N))$ と S_N は統計モデルとして等しく，幾何構造は同じである．

(E) ここで，$F \circ \Phi = I$ が成り立つ F について，$F(\cdot)$ が十分統計量であると言う．したがって，Φ がマルコフ埋め込みであれば，必ず $F \circ \Phi = I$ が成り立つ F が存在するので，$F(\cdot)$ という十分統計量が存在する．

(F) したがって，したがって，Φ がマルコフ埋め込みであれば，必ず $\Phi(S_N) = S_N$ が言える．すなわち，S_N と，これを S_M に埋め込んだ部分多様体 $\Phi(S_N)$ の幾何構造は等しい．

S_N の幾何構造は S_N の計量 g_N によって表現される．一方，S_M に，写像 Φ によって埋め込まれた部分多様体 $\Phi(S_N)$ の幾何構造は S_M の計量 g_M を写像 Φ で引き戻したもので表される． これを $(\Phi^* g_M)$ と書くことにすれば，Φ がマルコフ埋め込みであれば，任意の $p \in S_N$ について

$$(\Phi^* g_M)_p = (g_N)_p$$

が成り立つ．ここで，$\boldsymbol{v}$, $\boldsymbol{w} \in T_p N$ に対して，

$$(\Phi^* g_M)_p(\boldsymbol{v}, \boldsymbol{w}) = (g_M)_{\Phi(p)}(d\Phi_p(\boldsymbol{v}), d\Phi_p(\boldsymbol{w}))$$

であるので，

$$(g_N)_p(\boldsymbol{v}, \boldsymbol{w}) = (g_M)_{\Phi(p)}(d\Phi_p(\boldsymbol{v}), d\Phi_p(\boldsymbol{w})) \tag{5.14}$$

が成り立つ．次節（第 5.4 節）においては，式 (5.14) を用いてチェンツォフ (Chentsov) の定理を証明する．

余談

文献 [2] では,「十分統計量となる $F(\cdot)$ が存在することが，Φ がマルコフ埋め込みであることの定義である」としている．しかし，ここで十分統計量を持ち出すことで，全体の議論を分かりにくくしてしまっているように思う．[1] では十分統計量についての言及はほとんどなく，十分統計量を使わずマルコフ埋め込みを明示的に定義している．[1] の説明の仕方を拡張して,「マルコフ埋め込みであれば，その帰結として十分統計量となる $F(\cdot)$ が存在する」と議論する方が分かりやすい.

基本的には,「$\Phi(S_N)$ に対する十分統計量である $F(i)$ が存在すること」と,「写像 Φ がマルコフ埋め込みであること」は等価である．したがって，第 5.3 節で，[2] に従って，3 例のマルコフ埋め込みについて，それぞれが十分統計量を持つことをわざわざ証明しているが，これは単なる確認であり，議論の進行に必須の事ではない.（マルコフ埋め込みなら必ず十分統計量を持つからである.）

十分統計量とは何か： 我々が観測するのは，確率データ $y^1, \ldots, y^m$ であり，それを生成している確率モデルは通常未知である．この確率データに対して，十分統計量が見出せることは，このデータが，低次元の統計モデルから，マルコフ埋め込み写像によって埋め込まれた（見かけ上の）高次元の統計モデルから生成されていることを意味している．そのときの十分統計量の写像 $F(\cdot)$ はマルコフ埋め込み写像の逆像になっている．十分統計量が見出されれば,（つまり，マルコフ埋め込み写像の逆像を見出すことができれば）未知パラメータの推定を，より低次元のデータ空間 $x^1, \ldots, x^n$（$n \leq m$）から行うことができる.

5.4　チェンツォフ (Chentsov) の定理

5.4.1　$(0,2)$ 型テンソルに対するチェンツォフの定理

統計モデル S_{n-1} 上に計量（$(0,2)$ 型テンソル場）$g^{[n]}$ が，$S_{\ell-1}$ 上に計量（$(0,2)$ 型テンソル場）$g^{[\ell]}$ が，与えられているとする[4]．任意のマルコフ埋め込み $\Phi : S_{n-1} \to S_{\ell-1}$ に対して，

$$(\Phi^* g^{[\ell]})_p = g_p^{[n]} \tag{5.15}$$

[4] この節では，$S_{\ell-1}$ 上の計量を $g^{[\ell]}$，S_{n-1} 上の計量を $g^{[n]}$ と表す.

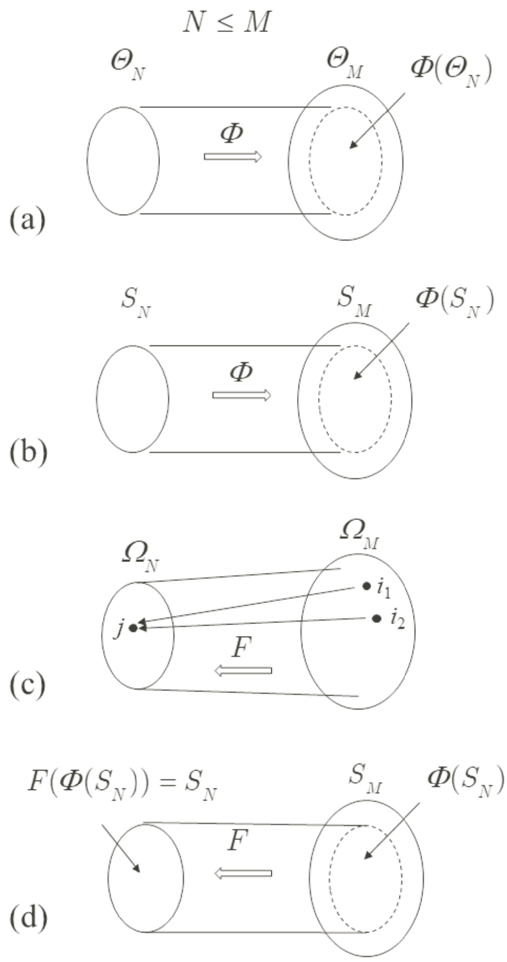

図 5.3: 多様体・パラメータ空間の間の写像とその向き

となるのは，$i, j = 1, \ldots, n$，$p = (x^1, \ldots, x^n) \in S_{n-1}$ に対して

$$g_{ij}^{[n]}(p) = g_p^{[n]}(\boldsymbol{e}_i, \boldsymbol{e}_j) = C\left(\frac{\delta_{ij}}{x_i} + \frac{1}{1 - \sum_{k=1}^{n} x_k}\right) \tag{5.16}$$

のときに限る．C はある実数（定数）である．

証明は，第 5.3 節で説明した例 1 から例 3 に表されるマルコフ埋め込み Φ に対して，式 (5.15) が成り立つなら，$g^{[n]}$ は式 (5.16) のように表される事を示す．つまり，必要条件を用いて，例１–例 3 の計量に対する制約条件を導出していく．

ここでは，文献 [1] の説明に沿った説明を行う．

式 (5.15) は，マルコフ埋め込みによって $S_{\ell-1}$ に埋め込まれた部分多様体の幾何構造が，もとの S_{n-1} の幾何構造に等しいことを表現したものである．以下で用いるために少し書き換えておく．まず，$\boldsymbol{v}, \boldsymbol{w} \in T_p S_{n-1}$ として，

$$(\Phi^* g^{[\ell]})_p(\boldsymbol{v}, \boldsymbol{w}) = g_{\Phi(p)}^{[\ell]}(d\Phi(\boldsymbol{v}), d\Phi(\boldsymbol{w})) = g_{\Phi(p)}^{[\ell]}(J_p(\Phi)(\boldsymbol{v}), J_p(\Phi)(\boldsymbol{w}))$$

であるので，$S_{\ell-1}$ に埋め込まれた部分多様体の幾何構造が，もとの S_{n-1} の幾何構造に等しいことの表現は，式 (5.14) を本節の表記に合わせて，

$$g_p^{[n]}(\boldsymbol{v},\boldsymbol{w})=g_{\Phi(p)}^{[\ell]}(J(\Phi)(\boldsymbol{v}),J(\Phi)(\boldsymbol{w})) \tag{5.17}$$

を得る．$J(\Phi)$ は写像の微分 $d\Phi$ を表すヤコビ行列である．

まず，以下の第 1 段階から第 3 段階までの議論では，マルコフ埋め込み Φ が，あたかも，

$$\Phi : \mathbb{R}_{++}^n \to \mathbb{R}_{++}^\ell \qquad (n \le \ell)$$

と定義されているとして議論を進める．議論の結果をを統計多様体 S_{n-1} へ制限するのは第 5.4.5 節で行う．

5.4.2　第 1 段階：例 1 における事象の並べ替え $\Phi : S_{n-1} \to S_{n-1}$ に対する不変性

まず，座標を S_{n-1} の重心 $p=(1/n.\ldots,1/n)\in\Omega_n$ にとる．Φ の微分 $d\Phi$ のヤコビ行列は

$$J(\Phi)=\begin{bmatrix}\boldsymbol{e}_{\zeta(1)}\\ \boldsymbol{e}_{\zeta(2)}\\ \vdots\\ \boldsymbol{e}_{\zeta(n)}\end{bmatrix}$$

であるので，誘導計量により $\boldsymbol{e}_i,\boldsymbol{e}_j\in T_PN(\Theta_n)$ の内積を計算する．

$$(\Phi^* g^{[n]})_p(\boldsymbol{e}_i,\boldsymbol{e}_j)=g_{\Phi(p)}^{[n]}\left(\boldsymbol{e}_iJ(\Phi),\boldsymbol{e}_jJ(\Phi)\right)$$

である．ここで，

$$\boldsymbol{e}_iJ(\Phi)=\boldsymbol{e}_{\zeta(i)} \qquad \boldsymbol{e}_jJ(\Phi)=\boldsymbol{e}_{\zeta(j)}$$

である．また，$p=(1/n,\ldots,1/n)$ を仮定しているので，

$$\Phi(p)=(1/n,\ldots,1/n)=p$$

である．したがって，

$$(\Phi^* g^{[n]})_p(\boldsymbol{e}_i,\boldsymbol{e}_j)=g_p^{[n]}\left(\boldsymbol{e}_{\zeta(i)},\boldsymbol{e}_{\zeta(j)}\right)$$

を得る．今，Φ は $\Theta_n\to\Theta_n$ の写像であるので，誘導計量は S_{n-1} の計量

$$g_p^{[n]}(\boldsymbol{e}_i,\boldsymbol{e}_j)$$

に等しい．つまり，式 (5.17) に対応して，

$$g_p^{[n]}(\boldsymbol{e}_i,\boldsymbol{e}_j)=g_p^{[n]}\left(\boldsymbol{e}_{\zeta(i)},\boldsymbol{e}_{\zeta(j)}\right) \tag{5.18}$$

が成り立たなければならない．したがって，計量 $g_p^{[n]}$ は i,j に依存しないことがわかる．(これは全事象が等確率となる重心点 $p=(1/n,\ldots,1/n)$ において，全事象が等確率で起こるなら事象の番号づけは意味がないという事実に対応する.)

したがって，この計量を，対角成分と非対角成分に分けて，

$$g_p^{[n]}(\boldsymbol{e}_i,\boldsymbol{e}_j)=\delta_{ij}A^{[n]}+B^{[n]}$$

とおく．つまり $g_p^{[n]}(\boldsymbol{e}_i, \boldsymbol{e}_j)$ の対角成分を $A^{[n]}$，非対角成分を $B^{[n]}$ とおいた．

ここで，S_{n-1} 上での制約

$$x_1 + x_2 + \cdots + x_n = 1 \tag{5.19}$$

を用いる[5]．S_{n-1} 上の接ベクトルの成分表示

$$\boldsymbol{X} = X^i \frac{\partial}{\partial x^i}$$

を式 (5.19) の両辺に作用させると，

$$\boldsymbol{X}\left(\sum_{k=1}^{n} x^k\right) = (X^i \frac{\partial}{\partial x^i})\left(\sum_{k=1}^{n} x^k\right) = \sum_{k=1}^{n} X^i \frac{\partial}{\partial x^i} x^k = \sum_{k=1}^{n} (X^i \delta_i^k) = \sum_{k=1}^{n} X^k = 0$$

を得る．したがって，任意の $\boldsymbol{X}, \boldsymbol{Y} \in T_p S_{n-1}$ に対して，

$$g_p^{[n]}(\boldsymbol{X}, \boldsymbol{Y}) = X^i Y^j g_p^{[n]}(\boldsymbol{e}_i, \boldsymbol{e}_j) = X^i Y^j (\delta_{ij} A^{[n]} + B^{[n]}) = A^{[n]} \sum_{i=1}^{n} X^i Y^i + B^{[n]} (\sum_{i=1}^{n} X^i)(\sum_{j=1}^{n} Y^j) = A^{[n]} \sum_{i=1}^{n} X^i Y^i$$

である．すなわち，$g_p^{[n]}(\boldsymbol{e}_i, \boldsymbol{e}_j)$ は $B^{[n]}$ に依存しない．したがって，一般性を失うことなく $B^{[n]} = 0$ とおいて，

$$g_p^{[n]}(\boldsymbol{e}_i, \boldsymbol{e}_j) = \delta_{ij} A^{[n]} \tag{5.20}$$

を得る．

5.4.3 第 2 段階：例 2 で用いたマルコフ埋め込み $\Phi : S_{n-1} \to S_{nm-1}$ に対する不変性

例 2 では，やはり，$p = (\frac{1}{n}, \ldots, \frac{1}{n}) \in \Omega_n$ を仮定する．したがって，

$$\Phi(p) = \Phi(x^1, \ldots, x^n) = \Big(\underbrace{\frac{x^1}{m}, \ldots, \frac{x^1}{m}}_{m\text{ 個}}, \underbrace{\frac{x^2}{m}, \ldots, \frac{x^2}{m}}_{m\text{ 個}}, \ldots, \underbrace{\frac{x^n}{m}, \ldots, \frac{x^n}{m}}_{m\text{ 個}}\Big) = \Big(\underbrace{\frac{1}{nm}, \ldots, \frac{1}{nm}}_{nm\text{ 個}}\Big)$$

である．Φ の微分 $d\Phi$ のヤコビ行列表現として，式 (5.10)：

$$d\Phi = J(\Phi) = \begin{bmatrix} \frac{1}{m}\sum_{\ell=1}^{m} \widetilde{\boldsymbol{e}}_\ell \\ \frac{1}{m}\sum_{\ell=m+1}^{2m} \widetilde{\boldsymbol{e}}_\ell \\ \vdots \\ \frac{1}{m}\sum_{\ell=(k-1)m+1}^{nm} \widetilde{\boldsymbol{e}}_\ell \end{bmatrix}$$

を用いて，誘導計量により $\boldsymbol{e}_i, \boldsymbol{e}_j \in T_P S_{n-1}$ の内積を計算する．式 (5.17) に対応して，

$$(\Phi^* g^{[nm]})_p(\boldsymbol{e}_i, \boldsymbol{e}_j) = (g^{[nm]})_{\Phi(p)}(\boldsymbol{e}_i J(\Phi), \boldsymbol{e}_j J(\Phi))$$

であり，式 (5.10) から，

$$\boldsymbol{e}_i J(\Phi) = (0, \ldots, 0, \underbrace{1}_{i\text{ 番目}}, 0, \ldots, 0) \begin{bmatrix} \frac{1}{m}\sum_{\ell=1}^{m} \widetilde{\boldsymbol{e}}_\ell \\ \frac{1}{m}\sum_{\ell=m+1}^{2m} \widetilde{\boldsymbol{e}}_\ell \\ \vdots \\ \frac{1}{m}\sum_{\ell=(k-1)m+1}^{nm} \widetilde{\boldsymbol{e}}_\ell \end{bmatrix} = \frac{1}{m} \sum_{\ell=(i-1)m+1}^{im} \widetilde{\boldsymbol{e}}_\ell$$

[5]苦肉の策で，第 5.4.5 節で行うはずの S_{n-1} への制限を部分的にここで行ってしまっている．

である．また，全く同じ理由で

$$\boldsymbol{e}_j J(\Phi) = \frac{1}{m} \sum_{\ell=(j-1)m+1}^{jm} \widetilde{\boldsymbol{e}}_\ell$$

である．したがって，

$$(\Phi^* g^{[nm]})_p(\boldsymbol{e}_i, \boldsymbol{e}_j) = g^{[nm]}_{\Phi(p)} \left(\frac{1}{m} \sum_{\ell=(i-1)m+1}^{im} \widetilde{\boldsymbol{e}}_\ell, \quad \frac{1}{m} \sum_{\ell'=(j-1)m+1}^{jm} \widetilde{\boldsymbol{e}}_{\ell'} \right) \tag{5.21}$$

となる．

第 1 段階の議論より $g^{[nm]}_{\Phi(p)}(\widetilde{\boldsymbol{e}}_\ell, \widetilde{\boldsymbol{e}}_{\ell'}) = A^{[nm]}\delta_{\ell\ell'}$ とおけば,（$\Phi(p) = (1/(nm), \ldots, 1/(nm))$ は全事象が等確率で起こる重心点なのでこの式が成立する.）式 (5.21) において，$i = j$ として，

$$\begin{aligned}
(\Phi^* g^{[nm]})_p(\boldsymbol{e}_i, \boldsymbol{e}_i) &= g^{[nm]}_{\Phi(p)} \left(\frac{1}{m} \sum_{\ell=(i-1)m+1}^{im} \widetilde{\boldsymbol{e}}_\ell, \quad \frac{1}{m} \sum_{\ell'=(i-1)m+1}^{im} \widetilde{\boldsymbol{e}}_{\ell'} \right) \\
&= (\frac{1}{m})^2 \sum_{\ell=(i-1)m+1}^{im} \sum_{\ell'=(i-1)m+1}^{im} g^{[nm]}_{\Phi(p)}(\widetilde{\boldsymbol{e}}_\ell, \widetilde{\boldsymbol{e}}_{\ell'}) \\
&= (\frac{1}{m})^2 \sum_{\ell=(i-1)m+1}^{im} \sum_{\ell'=(i-1)m+1}^{im} A^{[nm]}\delta_{\ell\ell'} = (\frac{1}{m})^2 m A^{[nm]} = \frac{A^{[nm]}}{m}
\end{aligned}$$

となる．ところで,

$$g^{[n]}_p(\boldsymbol{e}_i, \boldsymbol{e}_i) = A^{[n]}$$

であるので，不変性の要請：

$$(\Phi^* g^{[nm]})_p(\boldsymbol{e}_i, \boldsymbol{e}_j) = g^{[n]}_p(\boldsymbol{e}_i, \boldsymbol{e}_j)$$

を考慮すれば,

$$\frac{A^{[nm]}}{m} = A^{[n]} \quad \rightarrow \quad \frac{A^{[nm]}}{nm} = \frac{A^{[n]}}{n}$$

が言える．つまり，$\frac{A^{[n]}}{n}$ は n には依存しないので，ある定数 C を用いて

$$\frac{A^{[n]}}{n} = C \quad \rightarrow \quad A^{[n]} = Cn$$

と表すことができる．

5.4.4　第 3 段階：例 3 で用いたマルコフ埋め込み $\Phi : S_{n-1} \to S_\ell$ に対する不変性

S_{n-1} 上の任意の有理点（有理数を座標に持つ点）を共通の分母を持った分数として，

$$p = (x^1, \ldots, x^n) = (\frac{m_1}{\ell}, \frac{m_2}{\ell}, \ldots, \frac{m_n}{\ell})$$

と表す．この点の埋め込まれた先は

$$\Phi(p) = \Phi(x^1, \ldots, x^n) = \Big(\underbrace{\frac{x^1}{m_1}, \ldots, \frac{x^1}{m_1}}_{m_1\text{個}}, \underbrace{\frac{x^2}{m_2}, \ldots, \frac{x^2}{m_2}}_{m_2\text{個}}, \ldots, \underbrace{\frac{x^n}{m_n}, \ldots, \frac{x^n}{m_n}}_{m_n\text{個}}\Big) = \Big(\underbrace{\frac{1}{\ell}, \ldots, \frac{1}{\ell}}_{\ell\text{個}}\Big)$$

である．すなわち，$\Phi(p)$ は $S_{\ell-1}$ の重心点である．

行ベクトルを $\widetilde{\boldsymbol{e}}_k$ として，$\widetilde{\boldsymbol{e}}_k$ を用いて表した写像の微分のヤコビ行列による表現：式 (5.12)

$$d\Phi = \begin{bmatrix} \frac{1}{m_1}\sum_{k=1}^{m_1}\widetilde{\boldsymbol{e}}_k \\ \frac{1}{m_2}\sum_{k=m_1+1}^{m_1+m_2}\widetilde{\boldsymbol{e}}_k \\ \vdots \\ \frac{1}{m_n}\sum_{k=m_1+\cdots+m_{n-1}+1}^{m_1+\cdots+m_n}\widetilde{\boldsymbol{e}}_k \end{bmatrix}$$

を用いる．

$\boldsymbol{e}_i, \boldsymbol{e}_j \in T_P S_{n-1}$ の内積を計算する．

$$(\Phi^* g^{[\ell]})_p(\boldsymbol{e}_i, \boldsymbol{e}_j) = g^{[\ell]}_{\Phi(p)}(\boldsymbol{e}_i J(\Phi), \boldsymbol{e}_j J(\Phi)) \tag{5.22}$$

であり，式 (5.12) から，

$$\boldsymbol{e}_i J(\Phi) = \frac{1}{m_i}\sum_{k=m_1+\cdots+m_{i-1}+1}^{m_1+\cdots+m_i}\widetilde{\boldsymbol{e}}_k$$

である．また，全く同様に

$$\boldsymbol{e}_j J(\Phi) = \frac{1}{m_j}\sum_{k'=m_1+\cdots+m_{j-1}+1}^{m_1+\cdots+m_j}\widetilde{\boldsymbol{e}}_{k'}$$

である．したがって式 (5.22) に代入して，

$$\begin{aligned} g^{[\ell]}_{\Phi(p)}(\boldsymbol{e}_i J(\Phi), \boldsymbol{e}_j J(\Phi)) &= g^{[\ell]}_{\Phi(p)}\left(\frac{1}{m_i}\sum_{k=m_1+\cdots+m_{i-1}+1}^{m_1+\cdots+m_i}\widetilde{\boldsymbol{e}}_k, \frac{1}{m_j}\sum_{k'=m_1+\cdots+m_{j-1}+1}^{m_1+\cdots+m_j}\widetilde{\boldsymbol{e}}_{k'}\right) \\ &= \frac{1}{m_i}\frac{1}{m_j}\sum_{k=m_1+\cdots+m_{i-1}+1}^{m_1+\cdots+m_i}\sum_{k'=m_1+\cdots+m_{j-1}+1}^{m_1+\cdots+m_j} g^{[\ell]}_{\Phi(p)}(\widetilde{\boldsymbol{e}}_k, \widetilde{\boldsymbol{e}}_{k'}) \\ &= \frac{1}{m_i}\frac{1}{m_j}\sum_{k=m_1+\cdots+m_{i-1}+1}^{m_1+\cdots+m_i}\sum_{k'=m_1+\cdots+m_{j-1}+1}^{m_1+\cdots+m_j} A^{[\ell]}\delta_{kk'} \end{aligned}$$

である．ここで，

$$g^{[\ell]}_{\Phi(p)}(\widetilde{\boldsymbol{e}}_k, \widetilde{\boldsymbol{e}}_{k'}) = A^{[\ell]}\delta_{kk'}$$

を用いた．($\Phi(p) = (1/\ell, \ldots, 1/\ell)$ は全事象が等確率で起こる重心点なので上式が成立する．）上式において，$i = j$ とすれば，

$$g^{[\ell]}_{\Phi(p)}(\boldsymbol{e}_i J(\Phi), \boldsymbol{e}_i J(\Phi)) = \frac{1}{m_i^2}\sum_{k=m_1+\cdots+m_{i-1}+1}^{m_1+\cdots+m_i}\sum_{k'=m_1+\cdots+m_{i-1}+1}^{m_1+\cdots+m_i} A^{[\ell]}\delta_{kk'} = \frac{1}{m_i}A^{[\ell]} = \frac{C\ell}{m_i}$$

を得る．したがって，

$$g^{[n]}_p(\boldsymbol{e}_i, \boldsymbol{e}_i) = g^{[\ell]}_{\Phi(p)}(\boldsymbol{e}_i J(\Phi), \boldsymbol{e}_i J(\Phi)) = \frac{C\ell}{m_i}$$

であり，さらに，

$$p(i) = \frac{m_i}{\ell}$$

であるので，結局，

$$g^{[n]}_p(\boldsymbol{e}_i, \boldsymbol{e}_j) = \delta_{ij}\frac{C\ell}{m_i} = \delta_{ij}\frac{C}{p(i)} \qquad \left(p(i) = \frac{m_i}{\ell}\right) \tag{5.23}$$

を得る，すなわち，不変性の要求を満たす計量は，S_{n-1} の全ての有理点 p で式 (5.23) で表される．すると，テンソル場は連続なので，結局全ての点 $p \in S_{n-1}$ で式 (5.23) が成立することが結論できる[6]

5.4.5　計量の S_{n-1} への制限

前節までの議論の帰結である式 (5.23) は Φ が

$$\Phi : \mathbb{R}^n_{++} \to \mathbb{R}^\ell_{++} \qquad (n \leq \ell)$$

と定義されているとして導かれた．したがって，この結果を統計多様体 S_{n-1} へ制限する必要がある．

S_{n-1} は $\mathbb{R}^n_{++}$ の部分空間であり，$\mathbb{R}^n_{++}$ の座標 $(x^1, x^2, \ldots, x^n)$ に対して，$x^1 + x^2 + \cdots + x^n = 1$ で表される超平面である．したがって，上で導いた計量をこの超平面に制約する．超平面上の点の座標を $(\theta^1, \theta^2, \ldots, \theta^{(n-1)})$ とすれば,

$$\begin{aligned}
x^1 &= \theta^1 \\
x^2 &= \theta^2 \\
&\vdots \\
x^{n-1} &= \theta^{n-1} \\
x^n &= 1 - \sum_{k=1}^{n-1} \theta^k
\end{aligned}$$

が超平面の座標 $(\theta^1, \theta^2, \ldots, \theta^{(n-1)})$ と $\mathbb{R}^n_{++}$ の座標 $(x^1, x^2, \ldots, x^n)$ を結ぶ関係式である．この関係式は

$$p(i) = x^i = \sum_{k=1}^{n-1} \theta^k \delta^i_k + \left(1 - \sum_{k=1}^{n-1} \theta^k\right) \delta^i_n \tag{5.24}$$

と 1 行で書くこともできる．

$\{x^i\} \to \{\theta^i\}$ に伴う，接空間の基底の変換は，$\nu = 1, \ldots, n-1$ として,

$$\frac{\partial}{\partial \theta^\nu} = \frac{\partial x^i}{\partial \theta^\nu} \frac{\partial}{\partial x^i}$$

であり,

$$\begin{aligned}
\frac{\partial x^\nu}{\partial \theta^\nu} &= 1 \\
\frac{\partial x^n}{\partial \theta^\nu} &= -1 \\
\frac{\partial x^i}{\partial \theta^\nu} &= 0 \quad \text{上記以外}
\end{aligned}$$

であるので,

$$\frac{\partial}{\partial \theta^\nu} = \frac{\partial x^i}{\partial \theta^\nu} \frac{\partial}{\partial x^i} = \frac{\partial}{\partial x^\nu} - \frac{\partial}{\partial x^n}$$

[6] 文献 [1] による．若干，納得感に欠けるが，これは言えるのでしょう．

である，したがって，$\mu,\nu=1,\ldots,n-1$ として，

$$g_p^{[n]}\left(\frac{\partial}{\partial\theta^\mu},\frac{\partial}{\partial\theta^\nu}\right)=g_p^{[n]}\left(\frac{\partial}{\partial x^\mu}-\frac{\partial}{\partial x^n},\frac{\partial}{\partial x^\nu}-\frac{\partial}{\partial x^n}\right)=g_p^{[n]}\left(\frac{\partial}{\partial x^\mu},\frac{\partial}{\partial x^\nu}\right)+g_p^{[n]}\left(\frac{\partial}{\partial x^n},\frac{\partial}{\partial x^n}\right)$$

である．ここで，$g_p^{[n]}\left(\frac{\partial}{\partial x^\mu},\frac{\partial}{\partial x^n}\right)=0$ を用いた．また，$g_p^{[n]}\left(\frac{\partial}{\partial x^\mu},\frac{\partial}{\partial x^\nu}\right)=0$（$\mu\neq\nu$）であるので，結局

$$g_p^{[n]}\left(\frac{\partial}{\partial x^\mu},\frac{\partial}{\partial x^\nu}\right)=C\frac{\delta_\mu^\nu}{p(\mu)}\quad \text{および}\quad g_p^{[n]}\left(\frac{\partial}{\partial x^n},\frac{\partial}{\partial x^n}\right)=C\frac{1}{p(n)}$$

と書けるので，計量は

$$g_p^{[n]}\left(\frac{\partial}{\partial\theta^\mu},\frac{\partial}{\partial\theta^\nu}\right)=C\left(\frac{\delta_\mu^\nu}{p(\mu)}+\frac{1}{p(n)}\right)\tag{5.25}$$

と表される．以上でチャンツォフの定理（式 (5.16)）が証明された．

5.4.6 フィシャー計量の導出

式 (5.25) からフィシャー計量を導くのはそれほど難しくない．

式 (5.25) の計量は以下の式で表される．

$$g_p^{[n]}\left(\frac{\partial}{\partial\theta^\mu},\frac{\partial}{\partial\theta^\nu}\right)=C\left(\frac{\delta_\mu^\nu}{p(\mu)}+\frac{1}{p(n)}\right)=C\sum_{i=1}^{n}\frac{(\delta_\mu^i-\delta_n^i)(\delta_\nu^i-\delta_n^i)}{p(i)}\tag{5.26}$$

証明は簡単である．和の項に $i=1,2,\ldots,n$ を代入していけばよい．

$$\begin{aligned}
\frac{(\delta_\mu^i-\delta_n^i)(\delta_\nu^i-\delta_n^i)}{p(i)}&=0 && i\neq\mu,\quad i\neq\nu,\quad \text{and}\quad i\neq n\\
\frac{(\delta_\mu^i-\delta_n^i)(\delta_\nu^i-\delta_n^i)}{p(i)}&=\frac{1}{p(\mu)} && i=\mu\\
\frac{(\delta_\mu^i-\delta_n^i)(\delta_\nu^i-\delta_n^i)}{p(i)}&=\frac{1}{p(n)} && i=n
\end{aligned}$$

であるので，式 (5.26) が言える．ここで，式 (5.24) から，

$$\frac{\partial}{\partial\theta^\mu}p(i)=\frac{\partial}{\partial\theta^\mu}\left(\sum_{k=1}^{n-1}\theta^k\delta_k^i+\left(1-\sum_{k=1}^{n-1}\theta^k\right)\delta_n^i\right)=\delta_\mu^i-\delta_n^i$$

であるので，結局，

$$g_p^{[n]}\left(\frac{\partial}{\partial\theta^\mu},\frac{\partial}{\partial\theta^\nu}\right)=C\sum_{i=1}^{n}\frac{(\delta_\mu^i-\delta_n^i)(\delta_\nu^i-\delta_n^i)}{p(i)}=C\sum_{i=1}^{n}\frac{(\frac{\partial}{\partial\theta^\mu}p(i))(\frac{\partial}{\partial\theta^\nu}p(i))}{p(i)}\tag{5.27}$$

を得る．あとは，公式

$$\frac{\partial}{\partial\theta^\mu}p(i)=p(i)\frac{\partial}{\partial\theta^\mu}\log p(i)$$

を用いれば，フィシャー計量

$$g_p^{[n]}\left(\frac{\partial}{\partial\theta^\mu},\frac{\partial}{\partial\theta^\nu}\right)=C\sum_{i=1}^{n}p(i)(\frac{\partial}{\partial\theta^\mu}\log p(i))(\frac{\partial}{\partial\theta^\nu}\log p(i))\tag{5.28}$$

を導くことができる．

5.5　$(0,3)$ 型テンソルに対する不変性

$(0,3)$ 型テンソル $S^{[n]}$ に対して，$\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}\in T_pS_{n-1}$ として，任意のマルコフ埋め込み $\Phi: S_{n-1}\to S_{\ell-1}$ に対する不変性

$$S_p^{[n]}(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z})=S_{\Phi(p)}^{[\ell]}(d\Phi(\boldsymbol{X}),d\Phi(\boldsymbol{Y}),d\Phi(\boldsymbol{Z}))$$

を満たすものは，定数倍を除いて

$$S_p^{[n]}(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z})=\sum_{i=1}^{n}p(i)(\boldsymbol{X}\log p(i))(\boldsymbol{Y}\log p(i))(\boldsymbol{Z}\log p(i))$$

に限られる．

5.5.1　証明-第 1 段階：例 1 における事象の並べ替え $\Phi: S_{n-1}\to S_{n-1}$ に対する不変性

点 $p=(1/n,1/n,\ldots,1/n)$ を仮定する．この点において，$(0,3)$ 型テンソル $S_p^{[n]}(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z})$ が以下の不変性を持つ．すなわち，

$$S_p^{[n]}(\boldsymbol{e}_i,\boldsymbol{e}_j,\boldsymbol{e}_k)=S_p^{[n]}\left(\boldsymbol{e}_{\zeta(i)},\boldsymbol{e}_{\zeta(j)},\boldsymbol{e}_{\zeta(k)}\right) \tag{5.29}$$

である．したがって，$(0,3)$ 型テンソル $S_p^{[n]}(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z})$ は i,j,k には依存しないので，

$$S_p^{[n]}(\boldsymbol{e}_i,\boldsymbol{e}_j,\boldsymbol{e}_k)=\delta_{ijk}A^{[n]}+\delta_{ij}B^{[n]}+\delta_{jk}C^{[n]}+\delta_{ki}D^{[n]}+E^{[n]}$$

と置くことができる．ここで，δ_{ijk} は $i=j=k$ の時のみ 1 で，他はゼロと言う記号である．

ここで，

$$\boldsymbol{X}=X^i\frac{\partial}{\partial x^i},\quad \boldsymbol{X}=X^i\frac{\partial}{\partial x^i},\quad \boldsymbol{Z}=Z^i\frac{\partial}{\partial x^i}$$

を

$$x^1+x^2+\cdots+x^n=1$$

に作用させると，

$$\boldsymbol{X}(x^1+x^2+\cdots+x^n)=\sum_{i=1}^{n}X^i=0$$

を得，同様に，

$$\sum_{i=1}^{n}Y^j=0,\qquad \sum_{k=1}^{n}Z^k=0$$

であるので，

$$\begin{aligned}S_p^{[n]}(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z})&=A^{[n]}\sum_{i=1}^{n}X^iY^iZ^i+B^{[n]}(\sum_{k=1}^{n}Z^k)(\sum_{i=1}^{n}X^iY^i)+C^{[n]}(\sum_{i=1}^{n}X^i)(\sum_{j=1}^{n}Y^jZ^j)\\&\quad+D^{[n]}(\sum_{j=1}^{n}Y^j)(\sum_{i=1}^{n}X^iZ^i)+E^{[n]}(\sum_{i=1}^{n}X^i)(\sum_{j=1}^{n}Y^j)(\sum_{k=1}^{n}Z^k)=A^{[n]}\sum_{i=1}^{n}X^iY^iZ^i\end{aligned}$$

となる．つまり，$B^{[n]},C^{[n]},D^{[n]},E^{[n]}$ には依存しない．したがって，一般性を失わずにこれらをゼロとする．

5.5.2 証明-第 2 段階：例 2 で用いたマルコフ埋め込み $\Phi : S_{n-1} \to S_{nm-1}$ に対する不変性

まず，

$$(\Phi^* S^{[nm]})_p(\boldsymbol{e}_i, \boldsymbol{e}_j, \boldsymbol{e}_k) = S^{[nm]}_{\Phi(p)} \left(\frac{1}{m} \sum_{\alpha=(i-1)m+1}^{im} \widetilde{\boldsymbol{e}}_\alpha, \quad \frac{1}{m} \sum_{\beta=(j-1)m+1}^{jm} \widetilde{\boldsymbol{e}}_\beta, \quad \frac{1}{m} \sum_{\gamma=(k-1)m+1}^{km} \widetilde{\boldsymbol{e}}_\gamma \right) \tag{5.30}$$

であるが，$i = j = k$ とすれば，

$$\begin{aligned}(\Phi^* S^{[nm]})_p(\boldsymbol{e}_i, \boldsymbol{e}_i, \boldsymbol{e}_i) &= S^{[nm]}_{\Phi(p)} \left(\frac{1}{m} \sum_{\alpha=(i-1)m+1}^{im} \widetilde{\boldsymbol{e}}_\alpha, \quad \frac{1}{m} \sum_{\beta=(i-1)m+1}^{im} \widetilde{\boldsymbol{e}}_\beta, \quad \frac{1}{m} \sum_{\gamma=(i-1)m+1}^{im} \widetilde{\boldsymbol{e}}_\gamma \right) \\ &= (\frac{1}{m})^3 \sum_{\alpha=(i-1)m+1}^{im} \sum_{\beta=(i-1)m+1}^{im} \sum_{\gamma=(i-1)m+1}^{im} S^{[nm]}_{\Phi(p)} (\widetilde{\boldsymbol{e}}_\alpha, \widetilde{\boldsymbol{e}}_\beta, \widetilde{\boldsymbol{e}}_\gamma) \\ &= (\frac{1}{m})^3 \sum_{\alpha=(i-1)m+1}^{im} \sum_{\beta=(i-1)m+1}^{im} \sum_{\gamma=(i-1)m+1}^{im} A^{[nm]} \delta_{\alpha\beta\gamma} = (\frac{1}{m})^3 m A^{[nm]} = \frac{A^{[nm]}}{m^2}\end{aligned}$$

となる[7]．ところで，

$$(\Phi^* S^{[nm]})_p(\boldsymbol{e}_i, \boldsymbol{e}_i, \boldsymbol{e}_i) = S^{[n]}_p(\boldsymbol{e}_i, \boldsymbol{e}_i, \boldsymbol{e}_i) = A^{[n]}$$

であるので，

$$A^{[n]} = \frac{A^{[nm]}}{m^2} \quad \text{すなわち} \quad \frac{A^{[n]}}{n^2} = \frac{A^{[nm]}}{(nm)^2}$$

が成り立つ．結局 $A^{[n]}$ は n^2 に比例する，つまり，定数 C を用いて

$$A^{[n]} = Cn^2$$

と表すことができる．

5.5.3 証明-第 3 段階：例 3 で用いたマルコフ埋め込み $\Phi : S_{n-1} \to S_\ell$ に対する不変性

S_{n-1} 上の任意の有理点（有理数を座標に持つ点）を共通の分母を持った分数として，

$$p = (x^1, \ldots, x^n) = (\frac{m_1}{\ell}, \frac{m_2}{\ell}, \ldots, \frac{m_n}{\ell})$$

と表す．また，前節とほとんど同じ導出を用いて，

$$S^{[\ell]}_{\Phi(p)}(\boldsymbol{e}_i J(\Phi), \boldsymbol{e}_i J(\Phi), \boldsymbol{e}_i J(\Phi)) = \frac{1}{m_i^3} \sum_{k=m_1+\cdots+m_{i-1}+1}^{m_1+\cdots+m_i} \sum_{k'=m_1+\cdots+m_{i-1}+1}^{m_1+\cdots+m_i} \sum_{k''=m_1+\cdots+m_{i-1}+1}^{m_1+\cdots+m_i} A^{[\ell]} \delta_{kk'k''} = \frac{1}{m_i^2} A^{[\ell]} = \frac{C\ell^2}{m_i^2}$$

を得る．したがって，

$$S^{[n]}_p(\boldsymbol{e}_i, \boldsymbol{e}_i, \boldsymbol{e}_i) = S^{[\ell]}_{\Phi(p)}(\boldsymbol{e}_i J(\Phi), \boldsymbol{e}_i J(\Phi), \boldsymbol{e}_i J(\Phi)) = C \left(\frac{\ell}{m_i} \right)^2$$

であり，さらに，

$$p(i) = \frac{m_i}{\ell}$$

であるので，結局，

$$S^{[n]}_p(\boldsymbol{e}_i, \boldsymbol{e}_j, \boldsymbol{e}_k) = \delta_{ijk} \frac{C}{p(i)^2} \tag{5.31}$$

を得る．上式は，全ての有理点で成り立つがテンソルの連続性より，全ての実数点で成り立つことが結論できる．

[7] α, β, γ が全部同じになるのは $\alpha = \beta = \gamma = 1$，$\alpha = \beta = \gamma = 2$ などと全部で m 通りである．

5.5.4 証明-第 4 段階：上記結果の S_{n-1} へ制限

式 (5.31) はマルコフ埋め込み Φ が

$$\Phi : \mathbb{R}^n_{++} \to \mathbb{R}^\ell_{++} \qquad (n \le \ell)$$

と定義されているとして導かれた．したがって，この結果を統計多様体 S_{n-1} へ制限する必要がある．

S_{n-1} は $\mathbb{R}^n_{++}$ の部分空間であり，$\mathbb{R}^n_{++}$ の座標 $(x^1, x^2, \dots, x^n)$ に対して，$x^1 + x^2 + \cdots + x^n = 1$ で表される超平面である．したがって，上で導いた計量をこの超平面に制約する．超平面上の点の座標を $(\theta^1, \theta^2, \dots, \theta^{(n-1)})$ とすれば，

$$\begin{aligned} x^1 &= \theta^1 \\ x^2 &= \theta^2 \\ &\vdots \\ x^{n-1} &= \theta^{n-1} \\ x^n &= 1 - \sum_{k=1}^{n-1} \theta^k \end{aligned}$$

が超平面の座標 $(\theta^1, \theta^2, \dots, \theta^{(n-1)})$ と $\mathbb{R}^n_{++}$ の座標 $(x^1, x^2, \dots, x^n)$ を結ぶ関係式である．この関係式は

$$p(i) = x^i = \sum_{k=1}^{n-1} \theta^k \delta^i_k + \left(1 - \sum_{k=1}^{n-1} \theta^k\right) \delta^i_n \tag{5.32}$$

と 1 行で書くこともできる．

$\{x^i\} \to \{\theta^i\}$ に伴う，接空間の基底の変換は，$\nu = 1, \dots, n-1$ として，

$$\frac{\partial}{\partial \theta^\nu} = \frac{\partial x^i}{\partial \theta^\nu} \frac{\partial}{\partial x^i}$$

であり，

$$\begin{aligned} \frac{\partial x^\nu}{\partial \theta^\nu} &= 1 \\ \frac{\partial x^n}{\partial \theta^\nu} &= -1 \\ \frac{\partial x^i}{\partial \theta^\nu} &= 0 \quad \text{上記以外} \end{aligned}$$

であるので，

$$\frac{\partial}{\partial \theta^\nu} = \frac{\partial x^i}{\partial \theta^\nu} \frac{\partial}{\partial x^i} = \frac{\partial}{\partial x^\nu} - \frac{\partial}{\partial x^n}$$

である，したがって，$\mu, \nu, \omega = 1, \dots, n-1$ として，

$$\begin{aligned} S^{[n]}_p \left(\frac{\partial}{\partial \theta^\mu}, \frac{\partial}{\partial \theta^\nu}, \frac{\partial}{\partial \theta^\omega} \right) &= S^{[n]}_p \left(\frac{\partial}{\partial x^\mu} - \frac{\partial}{\partial x^n}, \frac{\partial}{\partial x^\nu} - \frac{\partial}{\partial x^n}, \frac{\partial}{\partial x^\omega} - \frac{\partial}{\partial x^n} \right) \\ &= S^{[n]}_p \left(\frac{\partial}{\partial x^\mu}, \frac{\partial}{\partial x^\nu}, \frac{\partial}{\partial x^\omega} \right) - S^{[n]}_p \left(\frac{\partial}{\partial x^n}, \frac{\partial}{\partial x^n}, \frac{\partial}{\partial x^n} \right) \end{aligned}$$

である．ここで，

$$S_p^{[n]}\left(\frac{\partial}{\partial x^n},\frac{\partial}{\partial x^\nu},\frac{\partial}{\partial x^\omega}\right)=0,\quad S_p^{[n]}\left(\frac{\partial}{\partial x^\mu},\frac{\partial}{\partial x^n},\frac{\partial}{\partial x^\omega}\right)=0,\quad S_p^{[n]}\left(\frac{\partial}{\partial x^\mu},\frac{\partial}{\partial x^\nu},\frac{\partial}{\partial x^n}\right)=0$$

であることを用いた．また，$S_p^{[n]}\left(\frac{\partial}{\partial x^\mu},\frac{\partial}{\partial x^\nu},\frac{\partial}{\partial x^\omega}\right)=0$（$\mu=\nu=\omega$ 以外の場合）であり，

$$S_p^{[n]}\left(\frac{\partial}{\partial x^\mu},\frac{\partial}{\partial x^\nu},\frac{\partial}{\partial x^\omega}\right)=C\frac{\delta_{\mu\nu\omega}}{p(\mu)^2}\quad \text{および}\quad S_p^{[n]}\left(\frac{\partial}{\partial x^n},\frac{\partial}{\partial x^n},\frac{\partial}{\partial x^n}\right)=C\frac{1}{p(n)^2}$$

であるので，計量は

$$S_p^{[n]}\left(\frac{\partial}{\partial \theta^\mu},\frac{\partial}{\partial \theta^\nu},\frac{\partial}{\partial \theta^\omega}\right)=C\left(\frac{\delta_{\mu\nu\omega}}{p(\mu)^2}-\frac{1}{p(n)^2}\right) \tag{5.33}$$

と表される．

式 (5.33) のテンソルは以下の式で表される．

$$S_p^{[n]}\left(\frac{\partial}{\partial \theta^\mu},\frac{\partial}{\partial \theta^\nu},\frac{\partial}{\partial \theta^\omega}\right)=C\left(\frac{\delta_{\mu\nu\omega}}{p(\mu)^2}-\frac{1}{p(n)^2}\right)=C\sum_{i=1}^{n}\frac{(\delta_\mu^i-\delta_n^i)(\delta_\nu^i-\delta_n^i)(\delta_\omega^i-\delta_n^i)}{p(i)^2} \tag{5.34}$$

であり，ここで，式 (5.24) から，

$$\frac{\partial}{\partial \theta^\mu}p(i)=\frac{\partial}{\partial \theta^\mu}\left(\sum_{k=1}^{n-1}\theta^k\delta_k^i+\left(1-\sum_{k=1}^{n-1}\theta^k\right)\delta_n^i\right)=\delta_\mu^i-\delta_n^i$$

であるので，結局，

$$S_p^{[n]}\left(\frac{\partial}{\partial \theta^\mu},\frac{\partial}{\partial \theta^\nu},\frac{\partial}{\partial \theta^\omega}\right)=C\sum_{i=1}^{n}\frac{(\frac{\partial}{\partial \theta^\mu}p(i))(\frac{\partial}{\partial \theta^\nu}p(i))(\frac{\partial}{\partial \theta^\omega}p(i))}{p(i)^2}$$

となる．あとは，公式

$$\frac{\partial}{\partial \theta^\mu}p(i)=p(i)\frac{\partial}{\partial \theta^\mu}\log p(i)$$

を用いれば，

$$S_p^{[n]}\left(\frac{\partial}{\partial \theta^\mu},\frac{\partial}{\partial \theta^\nu},\frac{\partial}{\partial \theta^\omega}\right)=C\sum_{i=1}^{n}p(i)(\frac{\partial}{\partial \theta^\mu}\log p(i))(\frac{\partial}{\partial \theta^\nu}\log p(i))(\frac{\partial}{\partial \theta^\omega}\log p(i)) \tag{5.35}$$

を導くことができる．

すなわち，$(0,3)$ 型テンソル $S^{[n]}$ に対して，任意のマルコフ埋め込み $\Phi : S_{n-1}\to S_{\ell-1}$ に対する不変性

$$S_p^{[n]}(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z})=S_{\Phi(p)}^{[\ell]}(d\Phi(\boldsymbol{X}),d\Phi(\boldsymbol{Y}),d\Phi(\boldsymbol{Z}))$$

を満たすものは，定数倍を除いて（式 (5.35) に示す）

$$S_p^{[n]}(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z})=\sum_{i=1}^{n}p(i)(\boldsymbol{X}\log p(i))(\boldsymbol{Y}\log p(i))(\boldsymbol{Z}\log p(i))$$

に限られることが示された．

5.6　アルファ接続

5.6.1　アルファ接続の導入

X, Y, Z を多様体 M のベクトル場（$X, Y, Z \in \mathfrak{X}(M)$）として，$T(X, Y, X)$ を $(0, 3)$ 型のテンソル場とする．すなわち，$T(X, Y, X)$ は

$$T(X, Y, Z)\,;\,\mathfrak{X}(M) \times \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathbb{R}$$

である線形多重写像である．

$T(X, Y, Z)$ を用いると，任意のアフィン接続 ∇ から新しいアフィン接続 $\widetilde{\nabla}$ を

$$g(\widetilde{\nabla}_Y X, Z) = g(\nabla_Y X, Z) + T(X, Y, Z) \tag{5.36}$$

として作り出すことができる．上式の ∇ がアフィン接続であることはアフィン接続の満たすべき条件

(i) $\nabla_{\boldsymbol{X}+\boldsymbol{Y}}\boldsymbol{Z} = \nabla_{\boldsymbol{X}}\boldsymbol{Z} + \nabla_{\boldsymbol{Y}}\boldsymbol{Z}$

(ii) $\nabla_{f\boldsymbol{X}}\boldsymbol{Z} = f\nabla_{\boldsymbol{X}}\boldsymbol{Z}$

(iii) $\nabla_{\boldsymbol{X}}(\boldsymbol{Y} + \boldsymbol{Z}) = \nabla_{\boldsymbol{X}}\boldsymbol{Y} + \nabla_{\boldsymbol{X}}\boldsymbol{Z}$

(iv) $\nabla_{\boldsymbol{X}}(f\boldsymbol{Y}) = (\boldsymbol{X}f)\nabla_{\boldsymbol{X}}\boldsymbol{Y} + f\nabla_{\boldsymbol{X}}\boldsymbol{Y}$

を ∇ が満たすことを確かめればよい．$T(X, Y, X)$ が線形多重写像であることから，この事はほとんど自明である．

ここで，リーマン接続 ∇ に対して，新しいアフィン接続 $\widetilde{\nabla}$ を

$$g(\widetilde{\nabla}_Y X, Z) = g(\nabla_Y X, Z) - \frac{\alpha}{2}T(X, Y, Z) \tag{5.37}$$

として定義する．ここで，$T(X, Y, Z)$ は式 (5.35) で定義した $(0, 3)$ 型テンソルである．$\frac{\alpha}{2}$ は定数であり，$\alpha = 0$ の場合 $\widetilde{\nabla} = \nabla$ である．すなわち，$\widetilde{\nabla}$ はリーマン接続に一致する．式 (5.37) で定義される新しいアフィン接続はアルファ接続と呼ばれる．この事を反映して新しい接続 $\widetilde{\nabla}$ を $\nabla^{(\alpha)}$ と表すことにする．すなわち，

$$g(\nabla^{(\alpha)}_Y X, Z) = g(\nabla_Y X, Z) - \frac{\alpha}{2}T(X, Y, Z) \tag{5.38}$$

と書く．

5.6.2　アルファ接続の接続係数の導出

この節ではアルファ接続の接続係数を導出する．式 (5.38) を再掲すると

$$g(\nabla^{(\alpha)}_Y X, Z) = g(\nabla_Y X, Z) - \frac{\alpha}{2}T(X, Y, Z)$$

に対して，$Y = \partial_i$, $X = \partial_j$, $Z = \partial_k$ を代入すれば

$$g(\nabla_Y X, Z) = g(\nabla_{\partial_i}\partial_j, \partial_k) = g(\Gamma^{\ell}{}_{ij}\,\partial_\ell, \partial_k) = \Gamma^{\ell}{}_{ij}\,g(\partial_\ell, \partial_k) = \Gamma^{\ell}{}_{ij}\,g_{\ell k} = \Gamma_{ij,k} \tag{5.39}$$

$$g(\nabla^{(\alpha)}_Y X, Z) = g(\nabla^{(\alpha)}_{\partial_i}\partial_j, \partial_k) = \Gamma^{(\alpha)}_{ij,k} \tag{5.40}$$

である．ここで，$\Gamma_{ij,k}$ がリーマン接続 ∇ から計算される接続係数（第 1 種クリストッフェル記号），$\Gamma^{(\alpha)}_{ij,k}$ がアルファ接続 $\nabla^{(\alpha)}$ から計算される接続係数（第 1 種クリストッフェル記号）である．以下では，式 (5.38)，(5.39)，(5.40) から $\Gamma_{ij,k}$ と $\Gamma^{(\alpha)}_{ij,k}$ の関係を導出する．

まず，$\Gamma_{ij,k}$ は良く知られた公式：

$$\Gamma_{ij,k} = \frac{1}{2}\left(-g_{ij,k} + g_{jk,i} + g_{ki,j}\right) = \frac{1}{2}\left(-\partial_k g_{ij} + \partial_i g_{jk} + \partial_j g_{ki}\right) \tag{5.41}$$

がある．この式の計算のためには具体的な計量が必要である．ここで，計量にフィシャー計量

$$g_{ij} = \sum_{x\in\Omega}[\partial_i \log p(x|\boldsymbol{\theta})][\partial_j \log p(x|\boldsymbol{\theta})]p(x|\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})]$$

を導入し，$l_{\boldsymbol{\theta}} = \log p(x|\boldsymbol{\theta})$ と書いて式 (5.41) の右辺を計算する．g_{ij} に ∂_k を作用させて，

$$\begin{aligned}\partial_k g_{ij} = &\sum_{x\in\Omega}[\partial_k\partial_i \log p(x|\boldsymbol{\theta})][\partial_j \log p(x|\boldsymbol{\theta})]p(x|\boldsymbol{\theta}) + \sum_{x\in\Omega}[\partial_i \log p(x|\boldsymbol{\theta})][\partial_k\partial_j \log p(x|\boldsymbol{\theta})]p(x|\boldsymbol{\theta}) \\ &+ \sum_{x\in\Omega}[\partial_i \log p(x|\boldsymbol{\theta})][\partial_j \log p(x|\boldsymbol{\theta})](\partial_k p(x|\boldsymbol{\theta}))\end{aligned} \tag{5.42}$$

を得る．ここで（よく使う）式変形

$$\partial_k p(x|\boldsymbol{\theta}) = [\partial_k \log p(x|\boldsymbol{\theta})]p(x|\boldsymbol{\theta})$$

を用いると，

$$\begin{aligned}\partial_k g_{ij} &= \sum_{x\in\Omega}[\partial_k\partial_i \log p(x|\boldsymbol{\theta})][\partial_j \log p(x|\boldsymbol{\theta})]p(x|\boldsymbol{\theta}) + \sum_{x\in\Omega}[\partial_i \log p(x|\boldsymbol{\theta})][\partial_k\partial_j \log p(x|\boldsymbol{\theta})]p(x|\boldsymbol{\theta}) \\ &\quad + \sum_{x\in\Omega}[\partial_i \log p(x|\boldsymbol{\theta})][\partial_j \log p(x|\boldsymbol{\theta})][\partial_k \log p(x|\boldsymbol{\theta})]p(x|\boldsymbol{\theta}) \\ &= E_{\boldsymbol{\theta}}[(\partial_k\partial_i \log p(x|\boldsymbol{\theta}))(\partial_j \log p(x|\boldsymbol{\theta}))] + E_{\boldsymbol{\theta}}[(\partial_i \log p(x|\boldsymbol{\theta}))(\partial_k\partial_j \log p(x|\boldsymbol{\theta})] \\ &\quad + E_{\boldsymbol{\theta}}[(\partial_i \log p(x|\boldsymbol{\theta}))(\partial_j \log p(x|\boldsymbol{\theta}))(\partial_k \log p(x|\boldsymbol{\theta}))] \\ &= E_{\boldsymbol{\theta}}[(\partial_k\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_k\partial_j l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})]\end{aligned}$$

となる．

すると，

$$\begin{aligned}\partial_k g_{ij} &= E_{\boldsymbol{\theta}}[(\partial_k\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_k\partial_j l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})] \\ \partial_i g_{jk} &= E_{\boldsymbol{\theta}}[(\partial_i\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_j l_{\boldsymbol{\theta}})(\partial_i\partial_k l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})(\partial_i l_{\boldsymbol{\theta}})] \\ \partial_j g_{ki} &= E_{\boldsymbol{\theta}}[(\partial_j\partial_k l_{\boldsymbol{\theta}})(\partial_i l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_k l_{\boldsymbol{\theta}})(\partial_j\partial_i l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_k l_{\boldsymbol{\theta}})(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})]\end{aligned}$$

であるので，

$$\begin{aligned}2\Gamma_{ij,k} &= -\partial_k g_{ij} + \partial_i g_{jk} + \partial_j g_{ki} \\ &= -\left(E_{\boldsymbol{\theta}}[(\partial_k\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_k\partial_j l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})]\right) \\ &\quad + \left(E_{\boldsymbol{\theta}}[(\partial_i\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_j l_{\boldsymbol{\theta}})(\partial_i\partial_k l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})(\partial_i l_{\boldsymbol{\theta}})]\right) \\ &\quad + \left(E_{\boldsymbol{\theta}}[(\partial_j\partial_k l_{\boldsymbol{\theta}})(\partial_i l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_k l_{\boldsymbol{\theta}})(\partial_j\partial_i l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_k l_{\boldsymbol{\theta}})(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})]\right) \\ &= 2E_{\boldsymbol{\theta}}[(\partial_i\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}}) + (\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})]\end{aligned}$$

を得る．したがって，

$$\Gamma_{ij,k} = E_{\boldsymbol{\theta}}\left[(\partial_i\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}}) + \frac{1}{2}(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})\right] = E_{\boldsymbol{\theta}}\left[\left(\partial_i\partial_j l_{\boldsymbol{\theta}} + \frac{1}{2}(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})\right)(\partial_k l_{\boldsymbol{\theta}})\right]$$

である．

α 接続に関する接続係数を求めるのであるが，ここで $(0,3)$ 型のテンソル場として，

$$\begin{aligned} T(X,Y,Z) &= \sum_{x\in\Omega} p(x|\boldsymbol{\theta})(X\log p(x|\boldsymbol{\theta}))(Y\log p(x|\boldsymbol{\theta}))(Z\log p(x|\boldsymbol{\theta})) \\ &= E_{\boldsymbol{\theta}}[(X\log p(x|\boldsymbol{\theta}))(Y\log p(x|\boldsymbol{\theta}))(Z\log p(x|\boldsymbol{\theta}))] \end{aligned} \tag{5.43}$$

を導入する.（統計多様体の不変性の要求を満たす $(0,3)$ 型のテンソルは，上記のみであるのが理由である.）このとき，成分 T_{ijk} は

$$T_{ijk} = T(\partial_i,\partial_j,\partial_k) = E_{\boldsymbol{\theta}}\left[(\frac{\partial}{\partial\theta^i}\log p(x|\boldsymbol{\theta}))(\frac{\partial}{\partial\theta^j}\log p(x|\boldsymbol{\theta}))(\frac{\partial}{\partial\theta^k}\log p(x|\boldsymbol{\theta}))\right]$$

として求まる．

すると，結局 α 接続に関する接続係数．

$$\Gamma^{(\alpha)}_{ij,k} = \Gamma_{ij,k} - \frac{\alpha}{2}E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})] = E_{\boldsymbol{\theta}}\left[\left(\partial_i\partial_j l_{\boldsymbol{\theta}} + \frac{1-\alpha}{2}(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})\right)(\partial_k l_{\boldsymbol{\theta}})\right] \tag{5.44}$$

を得る．上式からすぐに，

$$\Gamma^{(\alpha)}_{ij,k} = \Gamma^{(\alpha)}_{ji,k}$$

が成り立つことがわかる．したがって，アルファ接続 $\nabla^{(\alpha)}$ は捩率がゼロである．

5.6.3　$\nabla^{(\alpha)}$ と $\nabla^{(-\alpha)}$ の双対性

統計多様体 M に対してフィシャー計量

$$g_{ij} = E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})]$$

を考え，接続 $\nabla^{(-\alpha)}$ と $\nabla^{(\alpha)}$ の関係を見る．式 (5.42) において既に示したように，

$$\partial_k g_{ij} = E_{\boldsymbol{\theta}}[(\partial_k\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_k\partial_j l_{\boldsymbol{\theta}})] + +E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})] \tag{5.45}$$

が成り立つ．ところで，

$$g(\nabla^{(\alpha)}_{\partial_k}\partial_i,\partial_j) = \Gamma^{(\alpha)}_{ki,j} \qquad g(\partial_i,\nabla^{(-\alpha)}_{\partial_k}\partial_j) = \Gamma^{(-\alpha)}_{kj,i}$$

であり，アルファ接続のクリストッフェル記号は式 (5.44):

$$\Gamma^{(\alpha)}_{ij,k} = E_{\boldsymbol{\theta}}\left[\left(\partial_i\partial_j l_{\boldsymbol{\theta}} + \frac{1-\alpha}{2}(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})\right)(\partial_k l_{\boldsymbol{\theta}})\right]$$

で与えられるので，結局，

$$\begin{aligned} g(\nabla^{(\alpha)}_{\partial_k}\partial_i,\partial_j) + g(\partial_i,\nabla^{(-\alpha)}_{\partial_k}\partial_j) &= \Gamma^{(\alpha)}_{ki,j} + \Gamma^{(-\alpha)}_{kj,i} \\ &= E_{\boldsymbol{\theta}}\left[\left(\partial_k\partial_i l_{\boldsymbol{\theta}} + \frac{1-\alpha}{2}(\partial_k l_{\boldsymbol{\theta}})(\partial_i l_{\boldsymbol{\theta}})\right)(\partial_j l_{\boldsymbol{\theta}})\right] + E_{\boldsymbol{\theta}}\left[\left(\partial_k\partial_j l_{\boldsymbol{\theta}} + \frac{1+\alpha}{2}(\partial_k l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})\right)(\partial_i l_{\boldsymbol{\theta}})\right] \\ &= E_{\boldsymbol{\theta}}[(\partial_k\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})] + E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_k\partial_j l_{\boldsymbol{\theta}})] + +E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})] = \partial_k g_{ij} \end{aligned}$$

が成り立つ．すなわち，$\nabla^{(-\alpha)}$ は $\nabla^{(\alpha)}$ の双対接続である．したがって，$(M, \nabla^{(\alpha)}, g)$ は統計多様体である．つまり，統計モデル M はフィシャー計量とアルファ接続を付与することにより統計多様体になる．

ここまでは，統計モデル M に対して特別な仮定をおいていない．したがって，統計モデル一般に成り立つ．しかし，フィシャー計量とアルファ接続は，有限離散な確率分布の統計モデル S_{n-1} におけるマルコフ埋め込みに対する不変性から導かれたものである．S_{n-1} 以外の統計モデルでは，そもそもフィシャー計量とアルファ接続を用いる根拠はどこに？

5.7 S_{n-1} の双対微分幾何

5.7.1 双対性の証明

ここまでの節では有限確率分布空間 S_{n-1} におけるマルコフ埋め込みに対する不変性の要請から，S_{n-1} 上に許容される計量と接続がフィシャー計量とアルファ接続に限られることを明らかにした．この節では，有限確率分布空間 S_{n-1} の多様体としての構造を調べる．

有限確率分布空間 S_{n-1} において次の命題が成り立つ．ここで，フィシャー計量を g，$\alpha-$ 接続を $\nabla^{(\alpha)}$ とする．

1. 各 $\alpha \in \mathbb{R}$ に対し，$\nabla^{(\alpha)}$ の捩率はゼロ．
2. 各 $\alpha \in \mathbb{R}$ に対し，$\nabla^{(\alpha)}$ と $\nabla^{(-\alpha)}$ は g に関し双対．
3. S_{n-1} は双対構造 $(g, \nabla^{(-1)}, \nabla^{(1)})$ に関して双対平坦.

命題 1：「$\nabla^{(\alpha)}$ の捩率はゼロ」

すでに第 5.6.2 節でテンソルの成分を用いた証明が与えてあるが，成分を用いない証明を以下に示す．ここで，リーマン接続を ∇ と書く．アルファ接続の定義より，

$$g(\nabla_X^{(\alpha)}\boldsymbol{Y}, \boldsymbol{Z}) = g(\nabla_X\boldsymbol{Y}, \boldsymbol{Z}) - \frac{\alpha}{2}T(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$$

である．同じように，

$$g(\nabla_Y^{(\alpha)}\boldsymbol{X}, \boldsymbol{Z}) = g(\nabla_Y\boldsymbol{X}, \boldsymbol{Z}) - \frac{\alpha}{2}T(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z})$$

であるので，

$$g(\nabla_X^{(\alpha)}\boldsymbol{Y} - \nabla_Y^{(\alpha)}\boldsymbol{X} - [\boldsymbol{X}, \boldsymbol{Y}], \boldsymbol{Z}) = g(\nabla_X\boldsymbol{Y} - \nabla_Y\boldsymbol{X} - [\boldsymbol{X}, \boldsymbol{Y}], \boldsymbol{Z}) - \frac{\alpha}{2}\left(T(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) - T(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z})\right)$$

∇ はリーマン接続であって捩率ゼロであるので，

$$g(\nabla_X\boldsymbol{Y} - \nabla_Y\boldsymbol{X} - [\boldsymbol{X}, \boldsymbol{Y}], \boldsymbol{Z}) = 0$$

$(0, 3)$ 型テンソルの定義より，

$$T(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) = T(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z})$$

であるので，結局，任意の $\boldsymbol{Z}$ に対して，

$$g(\nabla_X^{(\alpha)}\boldsymbol{Y} - \nabla_Y^{(\alpha)}\boldsymbol{X} - [\boldsymbol{X}, \boldsymbol{Y}], \boldsymbol{Z}) = 0$$

を得る．

命題 2：「$\nabla^{(\alpha)}$ と $\nabla^{(-\alpha)}$ は g に関し双対」

第 5.6.3 において成分を用いた証明を行っているので，ここでは成分を用いない証明を行う．

まず，

$$\boldsymbol{X}g(\boldsymbol{Y},\boldsymbol{Z}) = g(\nabla_X \boldsymbol{Y},\boldsymbol{Z}) + g(\boldsymbol{Y},\nabla_X \boldsymbol{Z}) \tag{5.46}$$

が成り立つ．これは，リーマン接続では内積に対して微分のライプニッツ則が成り立つことを言っているに過ぎない．アルファ接続の定義より，

$$g(\nabla_X \boldsymbol{Y},\boldsymbol{Z}) = g(\nabla_X^{(\alpha)} \boldsymbol{Y},\boldsymbol{Z}) + \frac{\alpha}{2}T(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}) \tag{5.47}$$

である．また，

$$g(\nabla_X^{(-\alpha)} \boldsymbol{Z},\boldsymbol{Y}) = g(\nabla_X \boldsymbol{Z},\boldsymbol{Y}) + \frac{\alpha}{2}T(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{Y})$$

において内積の順序を変えて

$$g(\boldsymbol{Y},\nabla_X^{(-\alpha)} \boldsymbol{Z}) = g(\boldsymbol{Y},\nabla_X \boldsymbol{Z}) + \frac{\alpha}{2}T(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{Y})$$

を得るので，したがって，

$$g(\boldsymbol{Y},\nabla_X \boldsymbol{Z}) = g(\boldsymbol{Y},\nabla_X^{(-\alpha)} \boldsymbol{Z}) - \frac{\alpha}{2}T(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{Y}) \tag{5.48}$$

となる．式 (5.47) と式 (5.48) を式 (5.46) に代入して，対称性 $T(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}) = T(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{Y})$ を用いれば，

$$\begin{aligned}
\boldsymbol{X}g(\boldsymbol{Y},\boldsymbol{Z}) &= g(\nabla_X \boldsymbol{Y},\boldsymbol{Z}) + g(\boldsymbol{Y},\nabla_X \boldsymbol{Z}) \\
&= g(\nabla_X^{(\alpha)} \boldsymbol{Y},\boldsymbol{Z}) + \frac{\alpha}{2}T(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}) + g(\boldsymbol{Y},\nabla_X^{(-\alpha)} \boldsymbol{Z}) - \frac{\alpha}{2}T(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{Y}) \\
&= g(\boldsymbol{Y},\nabla_X^{(-\alpha)} \boldsymbol{Z}) + g(\nabla_X^{(\alpha)} \boldsymbol{Y},\boldsymbol{Z}) + \frac{\alpha}{2}\left(T(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}) - T(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{Y})\right) = g(\boldsymbol{Y},\nabla_X^{(-\alpha)} \boldsymbol{Z}) + g(\nabla_X^{(\alpha)} \boldsymbol{Y},\boldsymbol{Z})
\end{aligned}$$

を得る．したがって，$\nabla^{(-\alpha)}$ と $\nabla^{(\alpha)}$ は互いに双対である．

$(0,3)$ 型テンソル $T(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z})$ の $\boldsymbol{X}$ と $\boldsymbol{Y}$ の対称性により命題 1 が成り立ち，$T(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z})$ の $\boldsymbol{Y}$ と $\boldsymbol{Z}$ の対称性により命題 2 が成り立つ．

命題 3：「S_{n-1} は双対構造 $(g,\nabla^{(-1)},\nabla^{(1)})$ に関して双対平坦」

S_{n-1} が接続 $\nabla^{(-1)}$ に対して，アフィン座標系を持つことを示す．

アルファ接続の接続係数は

$$\Gamma_{ij,k}^{(\alpha)} = E_{\boldsymbol{\theta}}\left[\left(\partial_i\partial_j l_{\boldsymbol{\theta}} + \frac{1-\alpha}{2}(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})\right)(\partial_k l_{\boldsymbol{\theta}})\right]$$

で与えられるので，$\alpha = -1$ の場合は

$$\Gamma_{ij,k}^{(-1)} = E_{\boldsymbol{\theta}}\left[\left(\partial_i\partial_j l_{\boldsymbol{\theta}} + (\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})\right)(\partial_k l_{\boldsymbol{\theta}})\right] = \sum_{i=1}^{n} p(i)\left[\frac{(\partial_i p(i))(\partial_j p(i))}{p(i)^2} + \partial_i \frac{\partial_j p(i)}{p(i)}\right](\partial_k l_{\boldsymbol{\theta}})$$

であるが，分数の微分より

$$\frac{(\partial_i p(i))(\partial_j p(i))}{p(i)^2} + \partial_i \frac{\partial_j p(i)}{p(i)} = \frac{\partial_i\partial_j p(i)}{p(i)}$$

であるので，これを代入すれば，

$$\Gamma^{(-1)}_{ij,k} = \sum_{i=1}^{n} p(i)\left[\frac{(\partial_i p(i))(\partial_j p(i))}{p(i)^2} + \partial_i \frac{\partial_j p(i)}{p(i)}\right](\partial_k l_{\boldsymbol{\theta}}) = \sum_{i=1}^{n} p(i)\left[\frac{\partial_i\partial_j p(i)}{p(i)}\right](\partial_k l_{\boldsymbol{\theta}}) = \sum_{i=1}^{n}(\partial_i\partial_j p(i))(\partial_k l_{\boldsymbol{\theta}})$$

を得る．

ここで，$p \in S_{n-1}$ に対して，S_{n-1} の元を

$$p(i) = \sum_{a=1}^{n-1} \xi^a \delta_a^i + \left(1 - \sum_{a=1}^{n-1} \xi^a\right)\delta_n^i \tag{5.49}$$

と表す．つまり，S_{n-1} に座標系 $(\xi^1, \dots, \xi^{n-1})$ を定める．すると，

$$\partial_i\partial_j p(i) = \frac{\partial}{\partial \xi^i}\frac{\partial}{\partial \xi^j} p(i) = 0$$

が成り立つので，したがって，この座標系において

$$\Gamma^{(-1)}_{ij,k} = 0$$

が成り立つ．すなわち，座標系 $(\xi^1, \dots, \xi^{n-1})$ は S_{n-1} の $\nabla^{(-1)}-$ アフィン座標系となっている．したがって，座標系 $(\xi^1, \dots, \xi^{n-1})$ において $\nabla^{(-1)}-$ 曲率はゼロとなる．さらに，$\nabla^{(1)}-$ 曲率もゼロとなる[8]．

ここで，式 (5.49) で表される確率分布の族を混合型分布族と呼ぶ．この式で定められる座標 ξ^a を混合座標系と呼ぶ．$\alpha = -1$ に対応する接続 $\nabla^{(-1)}$ を混合型接続と呼び，記号 $\nabla^{(m)}$ で表す．これに対して，$\alpha = 1$ に対応する接続 $\nabla^{(1)}$ を指数型接続と呼び，記号 $\nabla^{(e)}$ で表す．

5.7.2 $\nabla^{(e)}-$ 双対アフィン座標の導出

以後 S_{n-1} を S と書く．統計多様体 $(S, g, \nabla^{(e)}, \nabla^{(m)})$ について議論する．

まず，双対アフィン座標系を定める．式 (5.49) に示す ξ^a が $\Gamma^{(-1)}_{ij,k} = 0$ を与える $\nabla^{(-1)}-$ アフィン座標系であるので，ξ^a は $\eta-$ 座標系に相当する．そこで，改めて式 (5.49) を

$$p(\omega) = \sum_{i=1}^{n-1} \eta_i \delta_i(\omega) + \left(1 - \sum_{i=1}^{n-1} \eta_i\right)\delta_n(\omega)$$

と書く．そして，$\boldsymbol{\eta} = (\eta_1, \dots, \eta_{n-1})$ を $\nabla^{(m)}-$ アフィン座標系として導入する．この座標系では

$$\partial_i\partial_j p(\omega) = \frac{\partial}{\partial \eta_i}\frac{\partial}{\partial \eta_j} p(\omega) = 0$$

であるので，$\nabla^{(m)}$ の接続係数

$$\Gamma^{(m)}_{ij,k} = g(\nabla^{(-1)}_{\partial_i}\partial_j, \partial_k) = \sum_{\omega=1}^{n}(\partial_i\partial_j p(\omega))(\partial_k \log p(\omega))$$

はゼロとなる．

[8] 双対性の重要な性質–∇ 曲率 R がゼロであることと，∇^* 曲率 R^* がゼロであることは同値である–からこれが言える．以前のノート「指数型分布族の多様体と双対平坦性」の第 3 節参照のこと．

次に，η_i の双対アフィン座標系を求める．$\log p(\omega)$ を以下のように書き直す．

$$\log p(\omega) = \sum_{i=1}^{n-1}\left(\log\frac{p(i)}{p(n)}\right)\delta_i(\omega) + \log p(n)$$

この式が成り立つことは,

- $\omega = 1$ とすれば,

$$\log p(1) = \sum_{i=1}^{n-1}\left(\log\frac{p(i)}{p(n)}\right)\delta_i(1) + \log p(n) = \left(\log\frac{p(1)}{p(n)}\right)\delta_1(1) + \log p(n) = \log p(1)$$

- $\omega = n$ とすれば,

$$\log p(n) = \sum_{i=1}^{n-1}\left(\log\frac{p(i)}{p(n)}\right)\delta_i(n) + \log p(n) = \left(\log\frac{p(n)}{p(n)}\right)\delta_n(n) + \log p(n) = \log p(n)$$

である．したがって，全ての ω で成り立つことがすぐわかる．

ここで,

$$\theta^i = \log\frac{p(i)}{p(n)} \qquad (i = 1, \ldots, n-1)$$

とする．規格化条件 $\sum_{\omega=1}^{n} p(\omega) = 1$ より,

$$p(n) = \frac{1}{1 + \sum_{i=1}^{n-1} e^{\theta^i}}$$

が成り立つ．したがって，$p(\omega)$ は

$$\log p(\omega) = \sum_{i=1}^{n-1}\theta^i\delta_i(\omega) - \log\left(\frac{1}{1 + \sum_{i=1}^{n-1} e^{\theta^i}}\right)$$

となる．ここで,

$$\psi(\boldsymbol{\theta}) = \log\left(\frac{1}{1 + \sum_{i=1}^{n-1} e^{\theta^i}}\right)$$

とすれば，$p(\omega)$ は,

$$\log p(\omega) = \sum_{i=1}^{n-1}\theta^i\delta_i(\omega) - \psi(\boldsymbol{\theta}) \tag{5.50}$$

と，指数型分布族として表すことができる．ここで，$\boldsymbol{\theta} = (\theta^1, \ldots, \theta^{n-1})$ を S の座標系と見立てて接続係数を計算する．すると，$\alpha = 1$ とした，α 接続の係数は,

$$\Gamma^{(1)}_{ij,k} = E_{\boldsymbol{\theta}}\left[(\partial_i\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})\right]$$

であり，結局,

$$\Gamma^{(e)}_{ij,k} = \Gamma^{(1)}_{ij,k} = E_{\boldsymbol{\theta}}\left[(\partial_i\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})\right] = -(\partial_i\partial_j\psi)\sum_{\omega=1}^{n}\partial_k p(\omega) = 0$$

である．ここで，$\partial_i\partial_j \log p(\omega) = -\partial_i\partial_j\psi(\boldsymbol{\theta})$ を用いた．したがって，$\boldsymbol{\theta}$ も平坦な座標系であり $\nabla^{(e)}-$ アフィン座標系であることがわかる．有限離散な確率分布は，混合型確率分布族としても指数型確率分布族としても表すことができる．

5.7.3 θ と η が双対アフィン座標であることの確認

η の双対アフィン座標系であることのダメ押しとして，

$$g\left(\frac{\partial}{\partial\eta_i}, \frac{\partial}{\partial\theta^j}\right)$$

を計算し，これが δ_i^j に等しいことを確認する．

まず，フィシャー計量：

$$g(\boldsymbol{X}, \boldsymbol{Y}) = \sum_{\omega=1}^{n}(\boldsymbol{X}p(\omega))(\boldsymbol{Y}\log p(\omega))$$

を用いれば，

$$g\left(\frac{\partial}{\partial\eta_i}, \frac{\partial}{\partial\theta^j}\right) = \sum_{\omega=1}^{n}(\frac{\partial}{\partial\eta_i}p(\omega))(\frac{\partial}{\partial\theta^j}\log p(\omega))$$

である．さらに，

$$\frac{\partial}{\partial\eta_i}p(\omega) = \frac{\partial}{\partial\eta_i}\left(\sum_{i=1}^{n-1}\eta_i\delta_i(\omega) + \left(1 - \sum_{i=1}^{n-1}\eta_i\right)\delta_n(\omega)\right) = \delta_i(\omega) - \delta_n(\omega)$$

である．また，

$$\frac{\partial}{\partial\theta^j}\log p(\omega) = \frac{\partial}{\partial\theta^j}\left(\sum_{i=1}^{n-1}\theta^i\delta_i(\omega) - \psi(\boldsymbol{\theta})\right) = \delta_j(\omega) - \frac{\partial\psi(\boldsymbol{\theta})}{\partial\theta^j}$$

であるので，

$$g\left(\frac{\partial}{\partial\eta_i}, \frac{\partial}{\partial\theta^j}\right) = \sum_{\omega=1}^{n}(\delta_i(\omega) - \delta_n(\omega))(\delta_j(\omega) - \frac{\partial\psi(\boldsymbol{\theta})}{\partial\theta^j}) = \sum_{\omega=1}^{n}(\delta_i(\omega) - \delta_n(\omega))\delta_j(\omega) = \delta_i^j$$

を得る．($\sum_{\omega=1}^{n}(\delta_i(\omega) - \delta_n(\omega)) = 0$ を用いた．）つまり，$\boldsymbol{\theta}$ と $\boldsymbol{\eta}$ は双対アフィン座標系であることが確認できる．

5.7.4 ポテンシャルの計算

まず，$\boldsymbol{\theta}-$ 座標系のポテンシャルは $\psi(\boldsymbol{\theta})$

$$\psi(\boldsymbol{\theta}) = \log\left(\frac{1}{1+\sum_{i=1}^{n-1}e^{\theta^i}}\right)$$

である．実際，

$$\partial_i\psi(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta^i}\log\left(\frac{1}{1+\sum_{i=1}^{n-1}e^{\theta^i}}\right) = \frac{e^{\theta^i}}{1+\sum_{i=1}^{n-1}e^{\theta^i}} = p(n)e^{\theta^i} = p(i) = \eta_i$$

を得る．また，$\boldsymbol{\eta}-$ 座標系のポテンシャル φ は

$$\varphi(\boldsymbol{\eta}) = \theta^i\eta_i - \psi(\boldsymbol{\theta}) = \sum_{i=1}^{n-1}\left(\log\frac{p(i)}{p(n)}\right)p(i) + \log p(n)$$

である．ここで，

$$p(n) = \frac{1}{1+\sum_{i=1}^{n-1}e^{\theta^i}}$$

であるので,

$$\psi(\boldsymbol{\theta}) = \log\left(\frac{1}{1+\sum_{i=1}^{n-1} e^{\theta^i}}\right) = -\log p(n)$$

を用いた．したがって,

$$\varphi(\boldsymbol{\eta}) = \sum_{i=1}^{n-1}\left(\log\frac{p(i)}{p(n)}\right)p(i) + \log p(n) = \sum_{i=1}^{n-1} p(i)\log p(i) + \left(1-\sum_{i=1}^{n-1} p(i)\right)\log p(n)$$

となる．さらに,

$$\sum_{i=1}^{n} p(i) = 1 \quad \rightarrow \quad p(n) = 1 - \sum_{i=1}^{n-1} p(i)$$

であるので，$\eta_i = p(i)$ として,

$$\varphi(\boldsymbol{\eta}) = \sum_{i=1}^{n-1}\eta_i\log\eta_i + \left(1-\sum_{i=1}^{n-1}\eta_i\right)\log\left(1-\sum_{i=1}^{n-1}\eta_i\right)$$

を得る．

したがって，$\partial^i\varphi$ を計算してみると,

$$\partial^i\varphi = \frac{\partial\varphi}{\partial\eta_i} = \log\eta_i + 1 - \log\left(1-\sum_{i=1}^{n-1}\eta_i\right) - 1 = \log\frac{p(i)}{p(n)} = \theta^i$$

を得る．

5.7.5　まとめ

双対構造

有限離散確率モデル S_{n-1} において，マルコフ埋め込みに対する不変性から，計量はフィシャー計量

$$g_{ij} = \sum_{\omega\in\Omega} p(\omega)(\frac{\partial}{\partial\theta^i}\log p(\omega))(\frac{\partial}{\partial\theta^j}\log p(\omega)) = E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})] \qquad l_{\boldsymbol{\theta}} = \log p(\omega)$$

とアルファ接続

$$\Gamma^{(\alpha)}_{ij,k} = E_{\boldsymbol{\theta}}\left[\left(\partial_i\partial_j l_{\boldsymbol{\theta}} + \frac{1-\alpha}{2}(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})\right)(\partial_k l_{\boldsymbol{\theta}})\right]$$

が決まる．すると，$(S_{n-1}, g, \nabla^{(-\alpha)}, \nabla^{(\alpha)})$ は双対構造を持つ多様体である．

双対平坦性

任意の点 $p\in S_{n-1}$ は,

$$p(\omega) = \sum_{i=1}^{n-1}\eta_i\delta_i(\omega) + \left(1-\sum_{i=1}^{n-1}\eta_i\right)\delta_n(\omega)$$

と表すことができる．この表し方における座標系 η_i について，

$$\partial^i \partial^j p(\omega) = \frac{\partial}{\partial \eta_i} \frac{\partial}{\partial \eta_j} p(\omega) = 0$$

が成り立つ．ところで，アルファ接続において $\alpha = -1$ とした $\nabla^{(m)}-$ 接続は

$$\Gamma_{ij,k}^{(-1)} = \sum_{i=1}^{n} (\partial_i \partial_j p(i))(\partial_k l_{\boldsymbol{\theta}})$$

と表されるが，$\partial_i \partial_j p(i) = 0$ であるので，この η_i 座標系では

$$\Gamma_{ij,k}^{(-1)} = 0$$

となる．したがって，S_{n-1} は，$(S_{n-1}, g, \nabla^{(-1)}, \nabla^{(1)})$ は双対平坦な多様体となる．

$\nabla^{(e)}-$ アフィン座標系

S_{n-1} における $\nabla^{(m)}-$ アフィン座標系は η_i である．η_i の双対アフィン座標系 θ^i は

$$\theta^i = \log \frac{p(i)}{p(n)} \qquad (i = 1, \ldots, n-1)$$

として求まる．実際，この座標系を用いて $p(\omega)$ を指数型分布族として，

$$\log p(\omega) = \sum_{i=1}^{n-1} \theta^i \delta_i(\omega) - \psi(\boldsymbol{\theta})$$

と表すことができる．$\alpha = 1$ とした，α 接続の係数は，

$$\Gamma_{ij,k}^{(1)} = E_{\boldsymbol{\theta}} \left[(\partial_i \partial_j l_{\boldsymbol{\theta}}) (\partial_k l_{\boldsymbol{\theta}}) \right]$$

である．上式の $\partial_i \partial_j l_{\boldsymbol{\theta}}$ は $\partial_i \partial_j l_{\boldsymbol{\theta}} = -\partial_i \partial_j \psi$ であるので，結局，θ^i 座標系を用いれば，

$$\Gamma_{ij,k}^{(1)} = E_{\boldsymbol{\theta}} \left[(\partial_i \partial_j l_{\boldsymbol{\theta}}) (\partial_k l_{\boldsymbol{\theta}}) \right] = - (\partial_i \partial_j \psi) E_{\boldsymbol{\theta}} \left[(\partial_k l_{\boldsymbol{\theta}}) \right] = 0$$

が成り立つ．θ^i と η_i 座標系の間の直交関係

$$g\left(\frac{\partial}{\partial \eta_i}, \frac{\partial}{\partial \theta^j} \right) = \delta_i^j$$

を示すこともできる．

ポテンシャル関数

$e-$ 座標系のポテンシャルは $\psi(\boldsymbol{\theta})$

$$\psi(\boldsymbol{\theta}) = \log \left(\frac{1}{1 + \sum_{i=1}^{n-1} e^{\theta^i}} \right)$$

であり，実際,

$$\partial_i \psi(\boldsymbol{\theta}) = \frac{\partial}{\partial \theta^i} \log \left(\frac{1}{1 + \sum_{i=1}^{n-1} e^{\theta^i}} \right) = p(i) = \eta_i$$

となる．

また，$m-$ 座標系のポテンシャルは

$$\varphi(\boldsymbol{\eta}) = \sum_{i=1}^{n-1} \eta_i \log \eta_i + \left(1 - \sum_{i=1}^{n-1} \eta_i \right) \log \left(1 - \sum_{i=1}^{n-1} \eta_i \right)$$

と表され，実際,

$$\partial^i \varphi = \frac{\partial \varphi}{\partial \eta_i} = \log \eta_i + 1 - \log \left(1 - \sum_{i=1}^{n-1} \eta_i \right) - 1 = \log \frac{p(i)}{p(n)} = \theta^i$$

となる．

追記： 指数型分布族において，自然パラメータ θ^i に対する双対アフィンパラメータ η_i は，期待値座標系

$$\eta_i = E[F_i(\omega)]$$

と与えられる．これは，$p(\omega)$ を（$\eta_i = p(i)$ として）混合型分布

$$p(\omega) = \sum_{i=1}^{n-1} \eta_i \delta_i(\omega) + \left(1 - \sum_{i=1}^{n-1} \eta_i \right) \delta_n(\omega)$$

と表した場合の η_i と同じものである．

証明： $p(\omega)$ を指数型分布族として表すと,

$$\log p(\omega) = \sum_{i=1}^{n-1} \theta^i \delta_i(\omega) - \psi(\boldsymbol{\theta})$$

となる．したがって，この指数型分布族に対する期待値座標系 ζ_i は

$$\zeta_i = E[\delta_i(\omega)] = \sum_{\omega=1}^{n-1} p(\omega) \delta_i(\omega) = p(i)$$

となる．したがって，$\eta_i = \zeta_i$ である．

第6章　統計的推定問題への応用：クラメル・ラオの不等式

2023-7-18

情報幾何の応用として，クラメル・ラオの不等式の導出を議論する．とは言っても，導出には，幾何学は特に必要としない．クラメル・ラオの不等式の導出の後，有効推定量について議論する．この有効推定量についての議論は，情報幾何的考え方が従来の統計学に勝ったひとつの例ではないかと思う．なお，EM アルゴリズムの幾何については，別の資料として取りまとめる．

6.1　クラメル・ラオの不等式

6.1.1　パラメータがスカラー（1 次元）の場合（統計学教科書の説明）

未知パラメータを θ（スカラー）として，確率変数 X の観測値を $x_1, x_2, \ldots, x_n$ とする．確率変数 X の確率分布を $f(x|\theta)$ とすれば，n 個の独立な観測データが従う確率分布は

$$p(x_1, \ldots, x_n|\theta) = f(x_1|\theta)f(x_2|\theta)\cdots f(x_n|\theta)$$

である．ここで，$x_1, \ldots, x_n$ と書くところを，x と書いてこれらを意味するとすれば，n 個の独立な観測データ $x_1, x_2, \ldots, x_n$ を得る確率は

$$p(x|\theta) = p(x_1, \ldots, x_n|\theta)$$

と表せる．ここで，対数尤度を

$$l(x|\theta) = \log p(x|\theta)$$

と置く．また，

$$U(x|\theta) = \frac{\partial}{\partial\theta} l(x|\theta) = \frac{\partial}{\partial\theta} \log p(x|\theta) = \frac{1}{p(x|\theta)} \frac{\partial}{\partial\theta} p(x|\theta)$$

はスコアー関数と呼ばれる．公式：

$$\begin{aligned} E[U(x|\theta)] = \int_\Omega U(x|\theta)p(x|\theta)dx = \int_\Omega p(x|\theta)\frac{\partial}{\partial\theta}\log p(x|\theta)dx = \int_\Omega p(x|\theta)\frac{1}{p(x|\theta)}\frac{\partial}{\partial\theta}p(x|\theta)dx \\ = \int_\Omega \frac{\partial}{\partial\theta}p(x|\theta)dx = \frac{\partial}{\partial\theta}\int_\Omega p(x|\theta)dx = 0 \quad (6.1) \end{aligned}$$

はよく使われる．また，以下の $I_n(\theta)$ をフィシャー情報量 $I_n(\theta)$ と呼ぶ．すなわち，

$$I_n(\theta) = E[U(x|\theta)^2] = E[(\frac{\partial}{\partial\theta}\log p(x|\theta))^2]$$

である．すると，$E[U(x|\theta)] = 0$ であるので，

$$V[U(x|\theta)] = E[U(x|\theta)^2] - (E[U(x|\theta)])^2 = E[U(x|\theta)^2] = I_n(\theta)$$

である．

ここで，パラメータ θ の推定解を $\hat{\theta}$ で表す．$\hat{\theta}$ は不偏推定解であるとすれば，

$$\theta = E[\hat{\theta}] = \int_\Omega \hat{\theta} p(x|\theta) dx$$

が成り立つ．上式の両辺を θ で微分すれば，

$$1 = \frac{\partial}{\partial \theta} \int_\Omega \hat{\theta}(x) p(x|\theta) dx = \int_\Omega \hat{\theta}(x) \frac{\partial}{\partial \theta} p(x|\theta) dx = \int_\Omega \hat{\theta}(x) U(x|\theta) p(x|\theta) dx$$

であり，つまり，

$$E[\hat{\theta}(x) U(x|\theta)] = 1$$

が成り立つ．一方で，$E[U(x|\theta)] = 0$ であるので，両辺を θ 倍して，

$$E[\theta U(x|\theta)] = 0$$

も成り立つ．したがって，

$$E[(\hat{\theta}(x) - \theta) U(x|\theta)] = 1$$

である．シュワルツの不等式 $E[A^2]E[B^2] \geq E[A \cdot B]^2$ に当てはめれば，

$$E[(\hat{\theta}(x) - \theta)^2] E[U(x|\theta)^2] \geq E[(\hat{\theta}(x) - \theta) U(x|\theta)]^2 = 1$$

すなわち，

$$V(\hat{\theta}) I_n(\theta) \geq 1 \quad \rightarrow \quad V(\hat{\theta}) \geq \frac{1}{I_n(\theta)}$$

が成り立つ．上式がクラメル・ラオの不等式として知られているもので，推定値の分散の下限を表している．すなわち，ある未知な量を推定する場合に，どこまでも正確な推定は行えず，精度に限界があることを示している．推定結果が，この不等式の下限を達成する推定量を有効推定量と呼ぶ．

6.1.2　パラメータが多変数（多次元）の場合

この節の説明は文献 [2] による．（この説明には，全く幾何学は登場しない．文献 [1] では，もう少し幾何学っぽい説明をしているが，ものすごくわかりにくい．）

未知パラメータをベクトルで表し $\boldsymbol{\theta} = (\theta^1, \ldots, \theta^m) \in \mathbb{R}^m$，$\boldsymbol{\theta}$ の取りえる値の集合を Θ_m，$\boldsymbol{\theta}$ の推定結果を $\hat{\boldsymbol{\theta}}$ とする．（特に断らない限り，これらは列ベクトルとする．）m 次元統計モデル

$$M = \{p(\boldsymbol{x}|\boldsymbol{\theta}) \,|\, \boldsymbol{\theta} \in \Theta_m\}$$

を考え，G を M 上のフィシャー情報行列とする．すると，任意の $\boldsymbol{\theta} \in \Theta_m$ に対して，

$$V(\hat{\boldsymbol{\theta}}) - \boldsymbol{G}^{-1}(\boldsymbol{\theta}) \geq 0$$

が成り立つ．ここで，

$$V(\hat{\boldsymbol{\theta}}) = E[(\hat{\boldsymbol{\theta}} - E[\hat{\boldsymbol{\theta}}])(\hat{\boldsymbol{\theta}} - E[\hat{\boldsymbol{\theta}}])^T]$$

であり，$\hat{\boldsymbol{\theta}}$ の共分散行列を表す.

証明：　不偏推定量であるので，

$$E[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$$

が成り立つ．また，スカラーの場合に対応して，

$$\partial_i l(\theta) = \frac{\partial}{\partial \theta^i} l(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{\partial}{\partial \theta^i} \log p(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{1}{p(\boldsymbol{x}|\boldsymbol{\theta})} \frac{\partial}{\partial \theta^i} p(\boldsymbol{x}|\boldsymbol{\theta})$$

であるので，

$$E[\partial_i l(\theta)] = E\left[\frac{\partial}{\partial \theta^i} l(\boldsymbol{x}|\boldsymbol{\theta})\right] = \int_\Omega \left[\frac{1}{p(\boldsymbol{x}|\boldsymbol{\theta})} \frac{\partial}{\partial \theta^i} p(\boldsymbol{x}|\boldsymbol{\theta})\right] p(\boldsymbol{x}|\boldsymbol{\theta}) d\boldsymbol{x} = \frac{\partial}{\partial \theta^i} \int_\Omega p(\boldsymbol{x}|\boldsymbol{\theta}) d\boldsymbol{x} = 0$$

が成り立つ．また，上式の両辺を $\boldsymbol{\theta}$ 倍しても

$$E[\boldsymbol{\theta} \partial_i l(\theta)] = 0 \tag{6.2}$$

が成り立つ．さらに，$E[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$ の両辺を θ^i で微分すれば，$\mathbb{R}^m$ の基本ベクトルを $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m$ として[1]，

$$\boldsymbol{e}_i = \frac{\partial}{\partial \theta^i} \int \hat{\boldsymbol{\theta}} p(\boldsymbol{x}|\boldsymbol{\theta}) d\boldsymbol{x} = \int \hat{\boldsymbol{\theta}} \frac{\partial}{\partial \theta^i} p(\boldsymbol{x}|\boldsymbol{\theta}) d\boldsymbol{x} = \int [\hat{\boldsymbol{\theta}} \frac{\partial}{\partial \theta^i} \log p(\boldsymbol{x}|\boldsymbol{\theta})] p(\boldsymbol{x}|\boldsymbol{\theta}) d\boldsymbol{x} = E[\hat{\boldsymbol{\theta}} \partial_i l(\boldsymbol{\theta})] \tag{6.3}$$

が成り立つ．式 (6.2) と式 (6.3) を組み合わせれば

$$\boldsymbol{e}_i = E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \partial_i l(\boldsymbol{\theta})] \tag{6.4}$$

を得る．

任意の,m 次元実数ベクトル $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m \in \mathbb{R}^m$ を考える．これらは列ベクトルとする．式 (6.2) と，不偏推定量であること $E[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$ を用いて，

$$\hat{\boldsymbol{\theta}} - \sum_{i=1}^m \boldsymbol{v}_i \partial_i l(\theta) - E\left[\hat{\boldsymbol{\theta}} - \sum_{i=1}^m \boldsymbol{v}_i \partial_i l(\theta)\right] = \hat{\boldsymbol{\theta}} - \sum_{i=1}^m \boldsymbol{v}_i \partial_i l(\theta) - \boldsymbol{\theta} - \sum_{i=1}^m \boldsymbol{v}_i E[\partial_i l(\theta)] = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} - \sum_{i=1}^m \boldsymbol{v}_i \partial_i l(\theta) \tag{6.5}$$

である．そして，共分散行列 $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma} = E\left[\left(\hat{\boldsymbol{\theta}} - \sum_{i=1}^m \boldsymbol{v}_i \partial_i l(\theta) - E\left[\hat{\boldsymbol{\theta}} - \sum_{i=1}^m \boldsymbol{v}_i \partial_i l(\theta)\right]\right)\left(\hat{\boldsymbol{\theta}} - \sum_{i=1}^m \boldsymbol{v}_i \partial_i l(\theta) - E\left[\hat{\boldsymbol{\theta}} - \sum_{i=1}^m \boldsymbol{v}_i \partial_i l(\theta)\right]\right)^T\right]$$

を計算する．すると，

$$\begin{aligned}
\boldsymbol{\Sigma} = E&\left[\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} - \sum_{i=1}^m \boldsymbol{v}_i \partial_i l(\theta)\right)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} - \sum_{i=1}^m \boldsymbol{v}_i \partial_i l(\theta)\right)^T\right] \\
&= E\left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\right] - E\left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\sum_{i=1}^m \boldsymbol{v}_i \partial_i l(\theta))^T\right] - E\left[(\sum_{i=1}^m \boldsymbol{v}_i \partial_i l(\theta))(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\right] \\
&\qquad + E\left[(\sum_{i=1}^m \boldsymbol{v}_i \partial_i l(\theta))(\sum_{i=1}^m \boldsymbol{v}_i \partial_i l(\theta))^T\right] \\
&= E\left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\right] - \sum_{i=1}^m \boldsymbol{e}_i \boldsymbol{v}_i^T - \sum_{i=1}^m \boldsymbol{v}_i \boldsymbol{e}_i^T + E\left[(\sum_{i=1}^m \boldsymbol{v}_i \partial_i l(\theta))(\sum_{i=1}^m \boldsymbol{v}_i \partial_i l(\theta))^T\right]
\end{aligned}$$

[1]例えば，$\boldsymbol{e}_1 = (1, 0, 0, \ldots, 0)$, $\boldsymbol{e}_m = (0, 0, 0, \ldots, 1)$ である．

である．ここで，式 (6.4) を用いて，

$$E\left[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\sum_{i=1}^{m}\boldsymbol{v}_i\partial_i l(\theta))^T\right]=\sum_{i=1}^{m}E\left[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\partial_i l(\theta)\right]\boldsymbol{v}_i^T=\sum_{i=1}^{m}\boldsymbol{e}_i\boldsymbol{v}_i^T$$

を用いた．また，

$$E\left[(\sum_{i=1}^{m}\boldsymbol{v}_i\partial_i l(\theta))(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\right]=\sum_{i=1}^{m}\boldsymbol{v}_i\boldsymbol{e}_i^T$$

も用いた．つまり，結局，

$$\boldsymbol{\Sigma}=E\left[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\right]-\sum_{i=1}^{m}\boldsymbol{e}_i\boldsymbol{v}_i^T-\sum_{i=1}^{m}\boldsymbol{v}_i\boldsymbol{e}_i^T+E\left[(\sum_{i=1}^{m}\boldsymbol{v}_i\partial_i l(\theta))(\sum_{i=1}^{m}\boldsymbol{v}_i\partial_i l(\theta))^T\right]$$

を得る．ここで，まず，

$$E\left[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\right]=V(\hat{\boldsymbol{\theta}})$$

である．また，

$$\boldsymbol{v}_i=\boldsymbol{G}^{-1}\boldsymbol{e}_i$$

とすると，

$$\sum_{i=1}^{m}\boldsymbol{e}_i\boldsymbol{v}_i^T=\sum_{i=1}^{m}\boldsymbol{e}_i\boldsymbol{e}_i^T\boldsymbol{G}^{-1}=\boldsymbol{G}^{-1}\sum_{i=1}^{m}\boldsymbol{e}_i\boldsymbol{e}_i^T=\boldsymbol{G}^{-1}$$

$$\sum_{i=1}^{m}\boldsymbol{v}_i\boldsymbol{e}_i^T=\sum_{i=1}^{m}\boldsymbol{G}^{-1}\boldsymbol{e}_i\boldsymbol{e}_i^T=\boldsymbol{G}^{-1}$$

また，

$$E\left[(\sum_{i=1}^{m}\boldsymbol{v}_i\partial_i l(\theta))(\sum_{i=1}^{m}\boldsymbol{v}_i\partial_i l(\theta))^T\right]=\sum_{i,j=1}^{m}\boldsymbol{v}_i\boldsymbol{v}_j^T E[(\partial_i l(\theta))(\partial_j l(\theta))]=\boldsymbol{G}^{-2}\sum_{i,j=1}^{m}\boldsymbol{e}_i\boldsymbol{e}_j^T G_{ij}=\boldsymbol{G}^{-1}$$

である．ここで，$\boldsymbol{e}_i\boldsymbol{e}_j^T$ は (i,j) 成分のみ 1 で他の要素は 0 である行列である．この行列に G_{ij} をかけてすべての (i,j) について足し合わせたら行列 $\boldsymbol{G}$ になる．

したがって，

$$\boldsymbol{\Sigma}=V(\hat{\boldsymbol{\theta}})-\boldsymbol{G}^{-1}$$

である．$\boldsymbol{\Sigma}$ は半正定値なので，$V(\hat{\boldsymbol{\theta}})-\boldsymbol{G}^{-1}$ も半正定値行列である．

6.1.3　有効推定量

$$V(\widehat{\boldsymbol{\xi}})=G^{-1}(\boldsymbol{\xi})$$

を達成するとき，$\widehat{\boldsymbol{\xi}}$ を統計モデル M の有効推定量と呼ぶ．

定理：　統計モデル M のパラメータ $\boldsymbol{\xi}$ に有効推定量が存在するための必要十分条件は M が指数型分布族であって，$\boldsymbol{\xi}$ がその期待値座標系であることである．

証明：　まず,「統計モデル M のパラメータ ξ に有効推定量が存在すれば，M が指数型分布族であって ξ がその期待値座標系である.」を示す.

推定量 $\widehat{\boldsymbol{\xi}}$ が有効推定量であることの必要十分条件は，前節の説明より,

$$\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi} - \sum_{i=1}^{m} \boldsymbol{G}^{-1}\boldsymbol{e}_i \partial_i l(\boldsymbol{\xi}) = 0 \tag{6.6}$$

が成り立つことである.（式中のベクトルは列ベクトルと仮定している.）$\boldsymbol{G}^{-1}$ の (i,j) 要素を G^{ij} と表せば,

$$\sum_{i=1}^{m} \boldsymbol{G}^{-1}\boldsymbol{e}_i \partial_i l(\boldsymbol{\xi}) = \sum_{i=1}^{m} \begin{bmatrix} G^{11} & \cdots & G^{1m} \\ \vdots & \ddots & \vdots \\ G^{m1} & \cdots & G^{mm} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \partial_i l(\boldsymbol{\xi}) \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} G^{1i}\partial_i l(\boldsymbol{\xi}) \\ \sum_{i=1}^{m} G^{2i}\partial_i l(\boldsymbol{\xi}) \\ \vdots \\ \sum_{i=1}^{m} G^{mi}\partial_i l(\boldsymbol{\xi}) \end{bmatrix}$$

である．式 (6.6) の第 j 成分を考えると,

$$\widehat{\xi}^j - \xi^j - \sum_{i=1}^{m} G^{ji}\partial_i l(\boldsymbol{\xi}) = 0$$

を得る．上下の重複する添え字にアインシュタインの規約を用いれば

$$\widehat{\xi^i}(x) - \xi^i = G^{ij}\partial_j \log p(x|\boldsymbol{\xi})$$

となる．この式は

$$\partial_j \log p(x|\boldsymbol{\xi}) = G_{ji}(\widehat{\xi^i}(x) - \xi^i)$$

と書き直せる．ここで，$\boldsymbol{\theta}$ と $\boldsymbol{\eta}$ が双対アフィン座標系であるとして，$\boldsymbol{\xi} = \boldsymbol{\eta}$ であるとすれば（つまり，$\boldsymbol{\xi}$ を双対アフィン座標系の期待値座標系 $\boldsymbol{\eta}$ に設定すれば）,

$$\frac{\partial}{\partial \eta_j} \log p(x|\boldsymbol{\eta}) = g^{ij}\widehat{\eta}_i(x) - g^{ij}\eta_i$$

である．上式の導出には

$$g^{ij} = g\left(\frac{\partial}{\partial \eta_i}, \frac{\partial}{\partial \eta_j}\right) = g\left(\frac{\partial}{\partial \theta_i}, \frac{\partial}{\partial \theta_j}\right) = G_{ij}$$

を用いた.

ここで,

$$g^{ij} = \frac{\partial \theta^i}{\partial \eta_j}$$

であり,

$$\frac{\partial \psi}{\partial \eta_j} = \frac{\partial \psi}{\partial \theta^k}\frac{\partial \theta^k}{\partial \eta_j} = g^{ji}\eta_i$$

である．したがって,

$$\frac{\partial}{\partial \eta_j} \log p(x|\boldsymbol{\eta}) = \frac{\partial \theta^i}{\partial \eta_j}\widehat{\eta}_i(x) - \frac{\partial \psi}{\partial \eta_j}$$

を得る．両辺を η_j で積分すれば，

$$\log p(x|\boldsymbol{\eta}) = C + \theta^i(\boldsymbol{\eta})\widehat{\eta}_i(x) - \psi(\boldsymbol{\theta}(\boldsymbol{\eta}))$$

を得る．上式は，$p(x|\boldsymbol{\eta})$ が指数関数族となっていることを示している．

次に，「M が指数型分布族であって，ξ がその期待値座標系であれば，ξ には有効推定量が存在する．」を示す．逆方向の証明も同じ様に行える．指数分布族を仮定すれば，

$$\log p(x|\boldsymbol{\theta}) = C(x) + \theta^i u_i(x) - \psi(\boldsymbol{\theta})$$

したがって，推定解は

$$\frac{\partial}{\partial\theta^i}\log p(x|\boldsymbol{\theta}) = \frac{\partial}{\partial\theta^i}(C(x) + \theta^i u_i(x) - \psi(\boldsymbol{\theta})) = u_i(x) - \frac{\partial\psi}{\partial\theta^i} = 0$$

から求まる．ここで，$\dfrac{\partial\psi}{\partial\theta^i} = \eta_i$ であるので，結局，推定解として

$$\widehat{\eta}_i = u_i(x)$$

を得る．つまり，$u_i(x)$ を $\boldsymbol{\eta}$ 座標系の η_i に対する推定結果 $\widehat{\eta}_i$ に等しいと考えれば，

$$\log p(x|\boldsymbol{\eta}) = C(x) + \theta^i(\boldsymbol{\eta})\widehat{\eta}_i - \psi(\boldsymbol{\theta}(\boldsymbol{\eta}))$$

となる．上式を η_i で微分して，

$$\frac{\partial}{\partial\eta_j}\log p(x|\boldsymbol{\eta}) = \frac{\partial\theta^i}{\partial\eta_j}\left(\widehat{\eta}_i - \frac{\partial\psi(\boldsymbol{\theta}(\boldsymbol{\eta}))}{\partial\theta^i}\right)$$

である．ここで，

$$\frac{\partial\theta^i}{\partial\eta_j} = g^{ij} \qquad \frac{\partial\psi(\boldsymbol{\theta}(\boldsymbol{\eta}))}{\partial\theta^i} = \eta_i$$

を代入すれば，

$$\frac{\partial}{\partial\eta_j}\log p(x|\boldsymbol{\eta}) = g^{ij}\left(\widehat{\eta}_i - \eta_i\right)$$

であり，これから容易に

$$\widehat{\eta}_i - \eta_i = G^{ij}\frac{\partial}{\partial\eta_j}\log p(x|\boldsymbol{\eta})$$

を導ける．上式は，推定解 $\widehat{\eta}_i$ がクラメル・ラオの下限を達成していること，つまり，$\widehat{\eta}_i$ が有効推定量である事を示している．

この定理は，指数分布族に対して期待値座標系を用いれば，必ず有効推定量が得られることを主張している．この定理に関する議論は，情報幾何のアプローチが，統計学における伝統的な考え方に対して優位性を出張できる数少ない例の１つかと思う．通常の統計学では，個々の推定に対して推定解が有効推定量であるかを議論するのが普通で，「こうすれば必ず有効推定解が得られる」という議論はめったにない．

6.1.4　スカラー正規分布での計算例

双対座標

スカラー（1 次元）正規分布

$$p(x|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

を，指数分布族の形式

$$p(x|\boldsymbol{\theta}) = \exp\left(\theta^i u_i(x) - \psi(\boldsymbol{\theta}) + c(x)\right)$$

で表せば,

$$\log p(x|\boldsymbol{\theta}) = \log\left[\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]\right] = \frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \left(\frac{\mu^2}{2\sigma^2} + \log(\sqrt{2\pi}\sigma)\right)$$

であるので,

$$c(x) = 0, \quad \theta^1 = \frac{\mu}{\sigma^2}, \quad u_1(x) = x, \quad \theta^2 = -\frac{1}{2\sigma^2}, \quad u_2(x) = x^2$$

であり,

$$\psi(\theta^1, \theta^2) = -\frac{(\theta^1)^2}{4\theta^2} + \frac{1}{2}\log\left(-\frac{\pi}{\theta^2}\right) = -\frac{(\theta^1)^2}{4\theta^2} - \frac{1}{2}\log\left(-\theta^2\right) \tag{6.7}$$

である．ここで，$\boldsymbol{\eta}$ 座標を求めると，式 (6.18) より，

$$\eta_1 = E[u_1(x)] = E[x]$$
$$\eta_2 = E[u_2(x)] = E[x^2]$$

である．つまり，η_1 は原点周りの 1 次モーメント，すなわち期待値 μ であり，η_2 は 2 次モーメントである[2].

N 個の独立な観測が得られる場合

ここで，N 個の独立なデータ $D = \{x_1, \ldots, x_N\}$ が観測されたとする．すると，データ D を得る確率分布を $p(D|\boldsymbol{\theta})$ とすれば,

$$\log p(D|\boldsymbol{\theta}) = \sum_{n=1}^{N}\log p(x_n|\boldsymbol{\theta}) = \theta^i\sum_{n=1}^{N}u_i(x_n) - N\psi(\boldsymbol{\theta})$$

を得る．上式を θ^1 で微分してゼロとおけば,

$$\sum_{n=1}^{N}u_1(x_n) - N\frac{\partial\psi(\boldsymbol{\theta})}{\partial\theta^1} = \sum_{n=1}^{N}u_1(x_n) - N\eta_1 = 0$$

したがって,

$$\widehat{\eta}_1 = \widehat{\mu} = \frac{1}{N}\sum_{n=1}^{N}u_1(x_n) = \frac{1}{N}\sum_{n=1}^{N}x_n$$

を得る．上式は，η_1 は原点周りの 1 次モーメント，すなわち期待値 μ の推定結果として算術平均 $\frac{1}{N}\sum_{n=1}^{N}x_n$ が得られることを示している．算術平均は，期待値 μ の最尤推定解であり，クラメル・ラオの下限を達成する最良の推定解であることは知られている．

また，θ^2 で微分してゼロとおけば,

$$\sum_{n=1}^{N}u_2(x_n) - N\frac{\partial\psi(\boldsymbol{\theta})}{\partial\theta^2} = \sum_{n=1}^{N}u_2(x_n) - N\eta_2 = 0$$

[2] ψ を微分しても同じ結果が得られることは，簡単に確認できる．

$$\eta_1 = \frac{\partial\psi}{\partial\theta^1} = -\frac{\theta^1}{2\theta^2} = -\frac{1}{2}\frac{\mu}{\sigma^2}(-2\sigma^2) = \mu = E[x] \tag{6.8}$$

$$\eta_2 = \frac{\partial\psi}{\partial\theta^2} = \frac{1}{4}\left(\frac{\theta^1}{\theta^2}\right)^2 + \frac{1}{2\theta^2} = \mu^2 + \sigma^2 = E[x^2] \tag{6.9}$$

である．

したがって,

$$\widehat{\eta}_2 = \frac{1}{N}\sum_{n=1}^{N} u_2(x_n) = \frac{1}{N}\sum_{n=1}^{N} x_n^2$$

を得る．$\widehat{\eta}_2$ は（原点周りの）2 次モーメント $E(x^2)$ の推定結果に対応している．

6.2　指数分布族に対する双対理論：復習

この節の内容は，前資料「指数型分布族の多様体と双対平坦性」から重要と思われる部分を抜粋し，指数型分布族における双対理論を復習したものである．

6.2.1　指数分布族に対する双対平坦性

指数分布族に対する双対理論を復習する．

確率分布族の多様体 M に対して，フィシャー計量

$$g_{ij} = E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})]$$

と，$(0,3)$ 型テンソル

$$T(X,Y,Z) = E_{\boldsymbol{\theta}}[(X\log p(x|\boldsymbol{\theta}))(Y\log p(x|\boldsymbol{\theta}))(Z\log p(x|\boldsymbol{\theta}))]$$

を導入し，アルファ接続 $\nabla^{(\alpha)}$ を

$$g(\nabla_Y^{(\alpha)}X, Z) = g(\nabla_Y X, Z) - \frac{\alpha}{2}T(X,Y,Z)$$

として定義すれば，アルファ接続による接続係数 $\Gamma_{ij,k}^{(\alpha)}$ は

$$\Gamma_{ij,k}^{(\alpha)} = \Gamma_{ij,k} - \frac{\alpha}{2}E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})] = E_{\boldsymbol{\theta}}\left[\left(\partial_i\partial_j l_{\boldsymbol{\theta}} + \frac{1-\alpha}{2}(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})\right)(\partial_k l_{\boldsymbol{\theta}})\right] \tag{6.10}$$

として求まる．ここで，$\Gamma_{ij,k}$ はフィシャー計量を用いたリーマン接続による接続係数である．この式からアルファ接続による接続係数は

$$\Gamma_{ij,k}^{(\alpha)} = \Gamma_{ji,k}^{(\alpha)}$$

を満たすので，捩率がゼロであることがわかる．また,

$$\Gamma_{ij,k}^{(\alpha)} + \Gamma_{ij,k}^{(-\alpha)} = \partial_k g_{ij}$$

を満たすので，$\nabla^{(\alpha)}$ と $\nabla^{(-\alpha)}$ は互いに双対接続の関係にある．したがって，多様体 $(M, \nabla^{(\alpha)}, g)$ は統計多様体となる．

指数分布族の確率分布の導入： 確率密度分布が

$$p(x|\boldsymbol{\theta}) = \exp\left(\theta^i u_i(x) - \psi(\boldsymbol{\theta}) + c(x)\right) \tag{6.11}$$

と表される確率分布の族を指数分布族と呼ぶ．確率分布のパラメータが $\boldsymbol{\theta}$ であり，x が確率変数である．$\boldsymbol{\theta}$ は指数分布族の自然なパラメータと呼ばれる．

指数分布族の多様体（統計モデル）M を

$$M = \{p(x|\boldsymbol{\theta}) \,|\, p(x|\boldsymbol{\theta}) = \exp\left(\theta^i u_i(x) - \psi(\boldsymbol{\theta}) + c(x)\right), \boldsymbol{\theta} = (\theta^1, \ldots, \theta^n) \in \Theta_n\}$$

と定義する．

双対平坦性の証明： 式 (6.11) の確率密度分布より，

$$\partial_j l_{\boldsymbol{\theta}} = \frac{\partial}{\partial \theta^j}\left(\theta^i u_i(x) - \psi(\boldsymbol{\theta}) + c(x)\right) = u_i(x) - \frac{\partial \psi(\boldsymbol{\theta})}{\partial \theta^j}$$

であり，さらに，

$$\partial_i \partial_j l_{\boldsymbol{\theta}} = \frac{\partial}{\partial \theta^i}\left(u_j(x) - \frac{\partial \psi(\boldsymbol{\theta})}{\partial \theta^j}\right) = -\frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}(\boldsymbol{\theta})$$

を得る．したがって，$\alpha = 1$ の場合のアルファ接続係数 $\Gamma^{(1)}_{ij,k}$ は,（式 (6.10) を参考にして）

$$\Gamma^{(1)}_{ij,k} = E_{\boldsymbol{\theta}}\left[(\partial_i \partial_j l_{\boldsymbol{\theta}})(\partial_k l_{\boldsymbol{\theta}})\right] = -E_{\boldsymbol{\theta}}\left[\left(\frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}(\boldsymbol{\theta})\right)(\partial_k l_{\boldsymbol{\theta}})\right] = -\frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}(\boldsymbol{\theta}) E_{\boldsymbol{\theta}}\left[(\partial_k l_{\boldsymbol{\theta}})\right]$$

であるが,

$$E_{\boldsymbol{\theta}}\left[(\partial_k l_{\boldsymbol{\theta}})\right] = \sum_{x \in \Omega}\left(\frac{\partial}{\partial \theta^k}\log p(x|\boldsymbol{\theta})\right)p(x|\boldsymbol{\theta}) = \sum_{x \in \Omega}\left(\frac{\partial}{\partial \theta^k}p(x|\boldsymbol{\theta})\right) = \frac{\partial}{\partial \theta^k}\sum_{x \in \Omega}p(x|\boldsymbol{\theta}) = \frac{\partial}{\partial \theta^k}1 = 0$$

であるので，$\Gamma^{(1)}_{ij,k} = 0$ となる．つまり，このとき $\boldsymbol{\theta} = (\theta^1, \ldots, \theta^n)$ はアフィン座標系であり，アルファ接続 $\nabla^{(\alpha=1)}$ は平坦である．アフィン座標系が存在するのでリーマン曲率テンソル $R^{(\alpha=1)}$ がゼロとなることがわかる．また，双対接続 $\nabla^{(\alpha=-1)}$ に対してもリーマン曲率テンソル $R^{(\alpha=-1)}$ はゼロとなるので，結局，指数分布族の多様体 M は双対平坦である[3].

まとめ：統計モデル M は，フィシャー計量 g_{ij} とアルファ接続を用いた双対構造 $\nabla^{(-\alpha)}$，$\nabla^{(\alpha)}$ のもとで統計多様体である．さらに，M を指数分布族の統計モデルとすれば，$\alpha = \pm 1$ において M は双対平坦となる．このとき指数分布族の自然なパラメータ $\boldsymbol{\theta} = (\theta^1, \ldots, \theta^n)$ がアフィン座標系となる．

6.2.2 双対アフィン座標とルジャンドル変換

平坦な多様体上では局所アフィン座標系（接続が恒等的にゼロとなる座標系）が存在する．指数分布族の統計モデル M においては，接続 $\nabla^{(-1)}$ と $\nabla^{(1)}$ について双対平坦であるので，局所アフィン座標系も 2 種類存在する．そこで，$(\theta^1, \ldots, \theta^n)$ を $\nabla^{(1)}$ に関する局所アフィン座標系，$(\eta_1, \ldots, \eta^n)$ を $\nabla^{(-1)}$ に関する局所アフィン座標系とすれば，これらは，M において各点の周りで

$$g\left(\frac{\partial}{\partial \theta^i}, \frac{\partial}{\partial \eta_j}\right) = g\left(\partial_i, \partial^j\right) = \delta_i^j \tag{6.12}$$

を満たす双対アフィン座標系となる．

[3] [3] には，アルファ接続に対するリーマン曲率テンソル $R^{(\alpha)}_{ijkl}$ とリーマン接続に対する曲率テンソル R_{ijkl} の間に

$$R^{(\alpha)}_{ijkl} = \frac{1-\alpha^2}{2}R_{ijkl}$$

の関係があることが述べられている.（証明は載っていない．めんどくさそう!）上式を用いれば $\alpha = \pm 1$ で $R^{(\alpha)}_{ijkl} = 0$ となることは簡単に示せる．

補題：双対アフィン座標系 $\{\theta^i, \eta_i\}$ に対する計量は

$$g_{ij} = g(\partial_i, \partial_j) = g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\theta^j}\right) \qquad g^{ij} = g(\partial^i, \partial^j) = g\left(\frac{\partial}{\partial\eta_i}, \frac{\partial}{\partial\eta_j}\right)$$

に等しい．計量に関して

$$g_{ij} = \frac{\partial\eta_j}{\partial\theta^i} = \frac{\partial\eta_i}{\partial\theta^j}$$
$$g^{ij} = \frac{\partial\theta^j}{\partial\eta_i} = \frac{\partial\theta^i}{\partial\eta_j}$$
$$g_{ij}g^{jk} = \delta_i^k$$

が成り立つ．

証明：

$$\partial_i = \frac{\partial}{\partial\theta^i} = \frac{\partial\eta_k}{\partial\theta^i}\frac{\partial}{\partial\eta_k} = \frac{\partial\eta_k}{\partial\theta^i}\partial^k$$

を用いれば

$$g_{ij} = g(\partial_i, \partial_j) = g(\frac{\partial\eta_k}{\partial\theta^i}\partial^k, \partial_j) = \frac{\partial\eta_k}{\partial\theta^i}g(\partial^k, \partial_j) = \frac{\partial\eta_k}{\partial\theta^i}\delta_j^k = \frac{\partial\eta_j}{\partial\theta^i} \tag{6.13}$$

が成り立つ．ここで，g_{ij} は座標変換 $(\theta^i) \to (\eta_j)$ の座標変換のヤコビ行列 $\left(\frac{\partial\eta_j}{\partial\theta^i}\right)$ である．また，計量は対称であるので（内積は順序によらないので）

$$g_{ij} = g_{ji} = \frac{\partial\eta_i}{\partial\theta^j}$$

も成り立つ．

また，

$$\partial^i = \frac{\partial}{\partial\eta_i} = \frac{\partial\theta^k}{\partial\eta_i}\frac{\partial}{\partial\theta^k} = \frac{\partial\theta^k}{\partial\eta_i}\partial_k$$

したがって，

$$g^{ij} = g(\partial^i, \partial^j) = \frac{\partial\theta^k}{\partial\eta_i}g(\partial_k, \partial^j) = \frac{\partial\theta^k}{\partial\eta_i}\delta_k^j = \frac{\partial\theta^j}{\partial\eta_i} \tag{6.14}$$

である．g^{ij} は座標変換 $(\eta_i) \to (\theta^j)$ のヤコビ行列 $\left(\frac{\partial\theta^j}{\partial\eta_i}\right)$ である．また，計量の対称性から

$$g^{ij} = g^{ji} = \frac{\partial\theta^i}{\partial\eta_j}$$

も成り立つ．最後に

$$g_{ij}g^{jk} = \frac{\partial\eta_j}{\partial\theta^i}\frac{\partial\theta^k}{\partial\eta_j} = \delta_i^k$$

は，ヤコビ行列とその逆行列を乗じているので単位行列になる．

定理 (i) 上の補題で述べたごとく n 次元指数関数族 $(M, g, \nabla^{(1)}, \nabla^{(-1)})$ は双対平坦な多様体であり，$\theta = (\theta^1, \ldots, \theta^n)$, $\eta = (\eta_1, \ldots, \eta_n)$ を M の双対アフィン座標系とすれば g_{ij} と g^{ij} は

$$g_{ij} = g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\theta^j}\right) \qquad g^{ij} = g\left(\frac{\partial}{\partial\eta_i}, \frac{\partial}{\partial\eta_j}\right)$$

として計算できる．すると，ある凸関数 $\psi(\theta)$ と $\varphi(\eta)$ が存在し，

$$g_{ij} = \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j} \qquad g_{ij} = \frac{\partial^2 \varphi}{\partial \eta_i \partial \eta_j}$$

と表され，また，

$$\theta^i = \frac{\partial \varphi}{\partial \eta_i} \qquad \eta_i = \frac{\partial \psi}{\partial \theta^i} \qquad \psi(\boldsymbol{\theta}) + \varphi(\boldsymbol{\eta}) = \theta^k \eta_k$$

が成り立つ．ここで，$\psi(\boldsymbol{\theta})$，$\varphi(\boldsymbol{\eta})$ はポテンシャルと呼ばれる．つまり，θ^i はポテンシャル $\varphi(\boldsymbol{\eta})$ の η_i に関する勾配として，η^j はポテンシャル $\psi(\boldsymbol{\theta})$ の θ_j に関する勾配として求まるのである．

証明：　関係式：

$$\frac{\partial \eta_i}{\partial \theta^j} = \frac{\partial \eta_j}{\partial \theta^i}$$

は連立微分方程式

$$\eta_i = \frac{\partial \psi}{\partial \theta^i} \qquad i = 1, \ldots, n$$

において，ポテンシャル関数 ψ（凸関数 ψ）が解として存在することの条件である．したがって，式 (6.13) を用いて，

$$g_{ij} = \frac{\partial \eta_i}{\partial \theta^j} = \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}$$

を得る．また，関係式：

$$\frac{\partial \theta^j}{\partial \eta_i} = \frac{\partial \theta^i}{\partial \eta_j}$$

は連立微分方程式

$$\theta^i = \frac{\partial \varphi}{\partial \eta_i} \qquad i = 1, \ldots, n$$

において，ポテンシャル関数 φ（凸関数 φ）が解として存在することの条件である．したがって，式 (6.14) を用いて

$$g^{ij} = \frac{\partial \theta^i}{\partial \eta_j} = \frac{\partial^2 \varphi}{\partial \eta_i \partial \eta_j}$$

を得る．

ここで，$\psi + \varphi - \theta^k \eta_k$ の全微分を計算すると

$$d\psi + d\varphi - (d\theta^k)\eta_k - \theta^k(d\eta_k) = \frac{\partial \psi}{\partial \theta^i} d\theta^i + \frac{\partial \varphi}{\partial \eta_i} d\eta_i - \eta_i d\theta^i - \theta^i d\eta_i = \eta_i d\theta^i + \theta^i d\eta_i - \eta_i d\theta^i - \theta^i d\eta_i = 0$$

したがって，$\psi + \varphi - \theta^k \eta_k$ は定数関数である．ポテンシャル関数には定数項の任意性があるので，ポテンシャル関数に定数を加えることにより

$$\psi + \varphi - \theta^k \eta_k = 0 \tag{6.15}$$

とすることができる．

定理 (ii) 指数分布族の多様体は $\alpha = 1$ において M は双対平坦となる．このとき $\boldsymbol{\theta} = (\theta^1, \ldots, \theta^n)$ がアフィン座標系である．式 (6.11) から

$$\psi(\boldsymbol{\theta}) = \log \int_\Omega \exp(\boldsymbol{\theta} \cdot \boldsymbol{u}(x))\, dx$$

となるので，この $\psi(\boldsymbol{\theta})$ を用いて，$\psi(\boldsymbol{\theta})$ の 2 階偏微分を計算してみれば,

$$\frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j} = E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})] = g_{ij} \tag{6.16}$$

である．

証明：

$$\log p(x|\boldsymbol{\theta}) = \theta^i u_i(x) - \psi(\boldsymbol{\theta}) + c(x)$$

の両辺を θ^i と θ^j で 2 回微分すれば，右辺で残るのは $\psi(\boldsymbol{\theta})$ の項だけである．すなわち，

$$\frac{\partial^2}{\partial\theta^i\partial\theta^j}\log p(x|\boldsymbol{\theta}) = -\frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}$$

である．したがって，

$$\frac{\partial^2\psi}{\partial\theta^i\partial\theta^j} = E_{\boldsymbol{\theta}}[\frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}] = -E_{\boldsymbol{\theta}}[\frac{\partial^2}{\partial\theta^i\partial\theta^j}\log p(x|\boldsymbol{\theta})] = E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})] = g_{ij}$$

である．

追記： ここで，等号：

$$-E_{\boldsymbol{\theta}}[\frac{\partial^2}{\partial\theta^i\partial\theta^j}\log p(x|\boldsymbol{\theta})] = E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})]$$

の成立について追加の説明を行う．まず，

$$E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})] = \sum_{x\in\Omega}\left(\frac{\partial}{\partial\theta^i}\log p(x|\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta^j}\log p(x|\boldsymbol{\theta})\right)p(x|\boldsymbol{\theta}) = \sum_{x\in\Omega}\left(\frac{\partial}{\partial\theta^i}\log p(x|\boldsymbol{\theta})\right)\frac{\partial}{\partial\theta^j}p(x|\boldsymbol{\theta}) \quad (6.17)$$

である．ところで，

$$E_{\boldsymbol{\theta}}\left[\frac{\partial}{\partial\theta^i}\log p(x|\boldsymbol{\theta})\right] = \sum_{x\in\Omega}\left(\frac{\partial}{\partial\theta^i}p(x|\boldsymbol{\theta})\right) = \frac{\partial}{\partial\theta^i}\sum_{x\in\Omega}p(x|\boldsymbol{\theta}) = 0$$

であるので，

$$\frac{\partial}{\partial\theta^j}E\left[\frac{\partial}{\partial\theta^i}\log p(x|\boldsymbol{\theta})\right] = \frac{\partial}{\partial\theta^j}\sum_{x\in\Omega}\left(\frac{\partial}{\partial\theta^i}\log p(x|\boldsymbol{\theta})\right)p(x|\boldsymbol{\theta}) = 0$$

である．したがって，

$$\begin{aligned}&\frac{\partial}{\partial\theta^j}\sum_{x\in\Omega}\left(\frac{\partial}{\partial\theta^i}\log p(x|\boldsymbol{\theta})\right)p(x|\boldsymbol{\theta})\\ &\qquad = \sum_{x\in\Omega}\left(\frac{\partial^2}{\partial\theta^j\partial\theta^i}\log p(x|\boldsymbol{\theta})\right)p(x|\boldsymbol{\theta}) + \sum_{x\in\Omega}\left(\frac{\partial}{\partial\theta^i}\log p(x|\boldsymbol{\theta})\right)\frac{\partial}{\partial\theta^j}p(x|\boldsymbol{\theta}) = 0\end{aligned}$$

であるので，結局，

$$\sum_{x\in\Omega}\left(\frac{\partial}{\partial\theta^i}\log p(x|\boldsymbol{\theta})\right)\frac{\partial}{\partial\theta^j}p(x|\boldsymbol{\theta}) = -\sum_{x\in\Omega}\left(\frac{\partial^2}{\partial\theta^j\partial\theta^i}\log p(x|\boldsymbol{\theta})\right)p(x|\boldsymbol{\theta}) = -E_{\boldsymbol{\theta}}\left[\frac{\partial^2}{\partial\theta^j\partial\theta^i}\log p(x|\boldsymbol{\theta})\right]$$

が成り立つ．式 (6.17) の最右辺に代入して，

$$E_{\boldsymbol{\theta}}[(\partial_i l_{\boldsymbol{\theta}})(\partial_j l_{\boldsymbol{\theta}})] = -E_{\boldsymbol{\theta}}\left[\frac{\partial^2}{\partial\theta^j\partial\theta^i}\log p(x|\boldsymbol{\theta})\right] = E_{\boldsymbol{\theta}}\left[\frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}\right] = \frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}$$

を得る．**追記終わり**

すなわち，式 (4.29) の右辺はフィシャー計量 g_{ij} に等しく，したがって，定理 (i) における双対アフィン座標 $\boldsymbol{\theta}$ に対するポテンシャル $\psi(\boldsymbol{\theta})$ は指数関数族の右辺に含まれている $\psi(\boldsymbol{\theta})$ に等しい.（等しいのでそもそも同じ記号 ψ を用いている訳である.）

もう片方の双対アフィン座標 η_i は

$$\eta_i = \frac{\partial \psi(\boldsymbol{\theta})}{\partial \theta^i} = \frac{\partial}{\partial \theta^i} \log \int \exp\left(\boldsymbol{\theta} \cdot \boldsymbol{u}(x)\right) dx = \frac{\frac{\partial}{\partial \theta^i} \int \exp\left(\boldsymbol{\theta} \cdot \boldsymbol{u}(x)\right) dx}{\int \exp\left(\boldsymbol{\theta} \cdot \boldsymbol{u}(x)\right) dx}$$

$$= \int u_i(x) \exp\left(\boldsymbol{\theta} \cdot \boldsymbol{u}(x) - \psi(\boldsymbol{\theta})\right) dx = \int u_i(x) p(x|\boldsymbol{\theta}) dx = E_{\boldsymbol{\theta}}[u_i(x)] \quad (6.18)$$

として求まる．η_i は $u_i(x)$ の期待値に等しい．η_i は，$\alpha = -1$ の場合のアルファ接続 $\nabla^{(-1)}$ に対するアフィン座標系である．$\eta_i = E_{\boldsymbol{\theta}}[u_i(x)]$ を期待値座標系と呼ぶ．

$\boldsymbol{\eta}$ に対するポテンシャル $\varphi(\boldsymbol{\eta})$ を求める．まず，

$$\log p(x|\boldsymbol{\theta}) = \theta^i u_i(x) - \psi(\boldsymbol{\theta}) + c(x)$$

の両辺の期待値を取れば，

$$E_{\boldsymbol{\theta}}[\log p(x|\boldsymbol{\theta})] = \theta^i E(u_i(x)) - \psi(\boldsymbol{\theta}) + c(x) = \theta^i \eta_i - \psi(\boldsymbol{\theta}) + c(x)$$

であるので，

$$\theta^i = \frac{\partial}{\partial \eta_i} E_{\boldsymbol{\theta}}[\log p(x|\boldsymbol{\theta})]$$

を得る．したがって，

$$\varphi(\eta) = E_{\boldsymbol{\theta}}[\log p(x|\boldsymbol{\theta}(\boldsymbol{\eta}))]$$

を得る．つまり，$\varphi(\eta)$ は確率のエントロピーの符号を変えたものになっている．ポテンシャルは定数の加算に対する任意性があるので，

$$\varphi(\eta) = E_{\boldsymbol{\theta}}[\log p(x|\boldsymbol{\theta}(\boldsymbol{\eta}))] - E_{\boldsymbol{\theta}}[c(x)] \quad (6.19)$$

とすることももちろん可能である．上式はダイバージェンスの導出に用いられる．

6.2.3　スカラー正規分布での計算例

スカラー正規分布における，前節の諸変数の計算例を示す．

スカラー（1 次元）正規分布

$$p(x|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

を，指数分布族の形式

$$p(x|\boldsymbol{\theta}) = \exp\left(\theta^i u_i(x) - \psi(\boldsymbol{\theta}) + c(x)\right)$$

で表せば，

$$\log p(x|\boldsymbol{\theta}) = \log\left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]\right] = \frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \left(\frac{\mu^2}{2\sigma^2} + \log(\sqrt{2\pi}\sigma)\right)$$

であるので，

$$c(x) = 0, \quad \theta^1 = \frac{\mu}{\sigma^2}, \quad u_1(x) = x, \quad \theta^2 = -\frac{1}{2\sigma^2}, \quad u_2(x) = x^2$$

であり，

$$\psi(\theta^1, \theta^2) = -\frac{(\theta^1)^2}{4\theta^2} + \frac{1}{2}\log\left(-\frac{\pi}{\theta^2}\right) = -\frac{(\theta^1)^2}{4\theta^2} - \frac{1}{2}\log\left(-\theta^2\right) \quad (6.20)$$

である．(定数項は無視した.)

ここで，$\boldsymbol{\eta}$ 座標を求めると，

$$\eta_1 = \frac{\partial\psi}{\partial\theta^1} = \mu$$

$$\eta_2 = \frac{\partial\psi}{\partial\theta^2} = \frac{1}{4}\left(\frac{\theta^1}{\theta^2}\right)^2 + \frac{1}{2\theta^2} = \mu^2 + \sigma^2$$

である．また，ポテンシャル $\varphi(\boldsymbol{\eta})$ は

$$\varphi(\boldsymbol{\eta}) = E_{\boldsymbol{\theta}}[\log p(x|\boldsymbol{\theta})] = E_{\boldsymbol{\theta}}[-\log(\sqrt{2\pi}\sigma) - \frac{(x-\mu)^2}{2\sigma^2}] = -\log\sigma - \frac{E_{\boldsymbol{\theta}}[(x-\mu)^2]}{2\sigma^2} = -\log\sigma - \frac{\sigma^2}{2\sigma^2} = -\log\sigma$$

ところで，

$$\sigma^2 = \eta_2 - \eta_1^2$$

であるので，

$$\varphi(\eta_1, \eta_2) = -\frac{1}{2}\log(\eta_2 - \eta_1^2)$$

を得る．ここで，ちなみに，

$$\frac{\partial\varphi}{\partial\eta_1} = -\frac{1}{2}\frac{-2\eta_1}{\eta_2 - \eta_1^2} = \frac{\mu}{\sigma^2} = \theta^1$$

$$\frac{\partial\varphi}{\partial\eta_2} = -\frac{1}{2}\frac{1}{\eta_2 - \eta_1^2} = -\frac{1}{\sigma^2} = \theta^2$$

を得る．

$$\sigma^2 = -\frac{1}{2\theta^2} \qquad \mu = -\frac{\theta^1}{2\theta^2}$$

となることを参考にして，計量を計算してみると，

$$g_{11} = \frac{\partial^2\psi}{\partial(\theta^1)^2} = -\frac{1}{2\theta^2} = \sigma^2$$

$$g_{12} = g_{21} = \frac{\partial^2\psi}{\partial\theta^2\partial\theta^1} = -\frac{1}{4}(2\theta^1)(-\frac{1}{(\theta^2)^2}) = \frac{1}{2}\frac{\theta^1}{(\theta^2)^2} = 2\mu\sigma^2$$

$$g_{22} = \frac{\partial^2\psi}{\partial\theta^2\partial\theta^2} = \frac{\partial}{\partial\theta^2}\left(\frac{(\theta^1)^2}{4(\theta^2)^2} + \frac{1}{2}\frac{1}{\theta^2}\right) = \frac{(\theta^1)^2}{4}\left(-2\frac{1}{(\theta^2)^3}\right) - \frac{1}{2}\frac{1}{(\theta^2)^2} = -\frac{1}{2(\theta^2)^2}\left[\frac{(\theta^1)^2}{\theta^2} + 1\right]$$

ここで，

$$\frac{1}{2(\theta^2)^2} = 2\sigma^4 \qquad \frac{(\theta^1)^2}{\theta^2} = -2\frac{\mu^2}{\sigma^2}$$

であるので，

$$g_{22} = -2\sigma^4\left(-2\frac{\mu^2}{\sigma^2} + 1\right) = 4\mu^2\sigma^2 - 2\sigma^4$$

となる．

第 7 章　統計的推定問題への応用：EM アルゴリズムの幾何

original：2023-7-18, revised：2023-7-22 and 7-31
前回の説明への補足の節を追加：2023-8-10

EM アルゴリズムの幾何学的解釈を議論する．

7.1　射影定理

7.1.1　KL ダイバージェンス

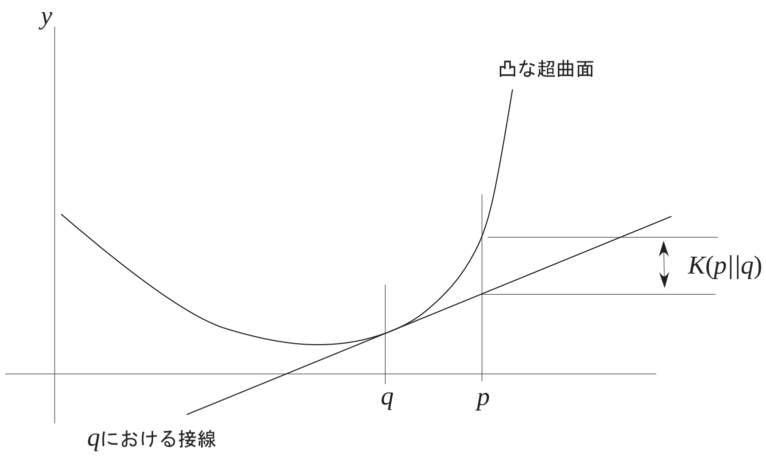

図 7.1: KL ダイバージェンス.

KL ダイバージェンスは以下のように定義される．

$$K(q\|p) = \int q(x)\log\frac{q(x)}{p(x)}dx = \int q(x)\left[\log q(x) - \log p(x)\right]dx$$

KL ダイバージェンスは，距離に近い性質を持っていて，偽距離と呼ばれることもある．,

以前のノートで議論したごとく，指数型分布族は（このノートでは特に断らない限り，指数型分布族を前提にする），双対平坦な多様体を構成し，ルジャンドル変換対をなす双対アフィン座標系を持つ．そして，これら座標系はポテンシャルと呼ばれるある凸関数の微分により与えられる．

ダイバージェンスとは，図 7.1 に示すように，ポテンシャルと呼ばれる凸関数の点 q における接平面と，凸関数が表す曲面との乖離の大きさを点 p で評価したものである．確率分布として指数型分布族を前提にするなら，このダイバージェンスは KL ダイバージェンスに等しくなる．

この図 7.1 から KL ダイバージェンスの大事な性質を見て取ることができる，まず，かならず $K(q\|p) \geq 0$ であること．次に，p と q が離れれば離れるほど $K(q\|p)$ は大きな値をとる．これは距離に近い性質である．しかしながら，$K(q\|p)$ の大きさは p と q の距離に比例するわけではないことも見て取れる．さらに，

$$K(q\|p) \neq K(p\|q)$$

であることがわかる．

7.1.2　KL ダイバージェンスとピタゴラスの定理

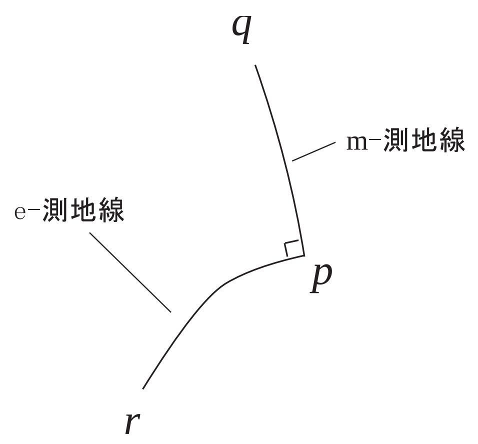

図 7.2: ピタゴラスの定理説明図

指数型分布族の確率分布の導入： 確率密度分布が

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \exp\left(\theta^i u_i(\boldsymbol{x}) - \psi(\boldsymbol{\theta}) + c(\boldsymbol{x})\right) \tag{7.1}$$

と表される確率分布の族を指数型分布族と呼ぶ．確率分布のパラメータが $\boldsymbol{\theta}$ であり，x が確率変数である．$\boldsymbol{\theta} = (\theta^1, \ldots, \theta^n)$ は指数分布族の自然なパラメータと呼ばれる．指数型分布族では，$E[u_1(\boldsymbol{x})], \ldots, E[u_n(\boldsymbol{x})]$ が，自然なパラメータ $\boldsymbol{\theta}$ に対する双対座標系を構成し，期待値座標系と呼ばれる．期待値座標系を

$$\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n) = (E[u_1(\boldsymbol{x})], \ldots, E[u_n(\boldsymbol{x})])$$

と記号 η_i で表す．

指数型分布族の空間の 2 点 $p(\boldsymbol{x})$ と $q(\boldsymbol{x})$ を考える．2 つの点の自然座標を $\boldsymbol{\theta}_p$ と $\boldsymbol{\theta}_q$ とする．2 点を結ぶ測地線は（この座標系では平坦なので）

$$\boldsymbol{\theta}(t) = (1-t)\boldsymbol{\theta}_p + t\boldsymbol{\theta}_q$$

と表すことができる．これを e-測地線と呼ぶ．同様に，2 つの点の期待値座標を $\boldsymbol{\eta}_p$ と $\boldsymbol{\eta}_q$ とする．2 点を結ぶ測地線は（この座標系も平坦なので）

$$\boldsymbol{\eta}(t) = (1-t)\boldsymbol{\eta}_p + t\boldsymbol{\eta}_q$$

と表すことができる．これを m-測地線と呼ぶ．

図 7.2 に示すように，指数型分布族の空間の 3 点 q, p, r を考える．q と p を結ぶ m-測地線が，p と r を結ぶ e-測地線に直交したとする．このとき，

$$K[q\|r] = K[q\|p] + K[p\|r]$$

が成り立つ．これは一般化されたピタゴラスの定理と呼ばれる．証明は以前のノート「指数型分布族の多様体と双対平坦性」参照のこと．

7.1.3 射影定理

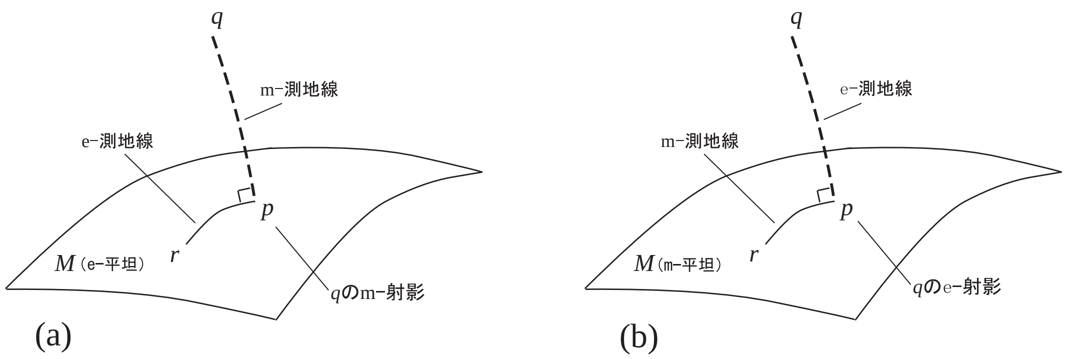

図 7.3: 射影定理.

指数型分布族の空間 S とその部分多様体 M が与えられたとする．S 上に点 q が与えられたとする．M の点で q 点に最も近い点を求める問題を考える．ここで，以下の射影定理が成り立つ．

1. 部分多様体 M 上の点 p と，その外にある点 q を結ぶ m-測地線が，p 点で M と直交するとき点 p を q から M への m 射影と言う（図 7.3 (a) 参照）．ここで，M を（親）多様体 S の $\nabla^{(e)}-$ 自己平行部分多様体と仮定する．すると，M は全測地的部分多様体でもあり，M の任意の 2 点間に引いた e-測地線は M に含まれる．これは，M も指数型分布族の多様体と仮定したことに等しい．当然ながら M は e-平坦である．さらに，q の M 上の m 射影 p は KL ダイバージェンス $K(q\|p)$ を最小にする．

2. 部分多様体 M 上の点 p と，その外にある点 q を結ぶ e-測地線が，p 点で M と直交するとき点 p を q から M への e 射影と言う（図 7.3 (b) 参照）．ここで，M を（親）多様体 S の $\nabla^{(m)}-$ 自己平行部分多様体と仮定する．すると，M は $\nabla^{(m)}-$ 全測地的部分多様体でもあり，M の任意の 2 点間に引いた m-測地線は M に含まれる．これは，M を混合型分布族の多様体と仮定した事に等しい．したがって，M は m-平坦である．さらに，q の M 上の e 射影 p は KL ダイバージェンス $K(p\|q)$ を最小にする．

証明： 1）の証明を行う．部分多様体 M の外に点 q を仮定し，p は q の（M への）m-射影とする．r を M 上の任意の点とする．(M は全測地的部分多様体であるので，r と p を結ぶ e-測地線は M に含まれ，(仮定より）この測地線は q と p を結ぶ m-測地線と p で直交する．したがって，ピタゴラスの定理より，

$$K[q\|r] = K[q\|p] + K[p\|r] \geq K[q\|p]$$

が成り立つ．つまり，$K[q\|r]$ は r が p に等しいときに最小になり，最小値 $K[q\|p]$ を持つ．

次に，２）の証明を行う．この場合，r と p を結ぶのが m-測地線で，p, q 間が e-測地線であるので，ピタゴラスの定理は

$$K[r\|q] = K[r\|p] + K[p\|q] \geq K[p\|q]$$

となる．したがって，$K[r\|q]$ は $r = p$ のとき最小で，最小値 $K[p\|q]$ を持つ．**証明終**

7.1.4 統計的推定

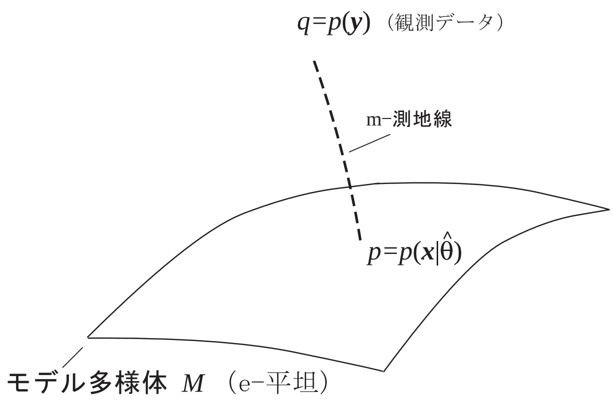

図 7.4: 推定の幾何学的解釈.

統計的な推定は S（S は指数型分布族とする）上のデータ点からモデルの空間 M への射影と解釈できる．今，データ $\boldsymbol{y}$ が観測されたとする．このデータを得る確率は $p(\boldsymbol{y})$ であり，S の中で点 $q = p(\boldsymbol{y})$ が観測データを表す点である．

一方のモデル多様体 M は S の e-平坦な部分多様体であると仮定する．これは M に対しても指数型分布族を仮定したことと等価である．

$$M = \{p(\boldsymbol{x}|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$$

とする．ここで，Θ はパラメータの定義域である．統計的推定とは，与えられた観測データから，観測データを生成した確率分布 $p(\boldsymbol{x}|\hat{\boldsymbol{\theta}})$ を推定する問題である．言い換えると，モデル多様体 M 上で，観測データを表す確率分布 $q = p(\boldsymbol{y})$ に最も近い点をもとめる問題である．

すると，図 7.4 に示すように，この推定は点 q を m-射影によりモデル M に射影することに等しい．なぜならば，ピタゴラスの定理から q の m-投影 p は KL ダイバージェンス $K[q\|p]$ を最小にする $p = p(\boldsymbol{x}|\boldsymbol{\theta})$ として求まる．したがって，パラメータの推定解は

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, K[p(\boldsymbol{y})\|p(\boldsymbol{x}|\boldsymbol{\theta})]$$

として求まる．

つまり，幾何学的には $q = p(\boldsymbol{y})$ から M へ下した m-測地線（垂線）の足 p を求める．点 p に対応した確率分布が，$p = p(\boldsymbol{x}|\hat{\boldsymbol{\theta}})$ である.（図 7.4 参照のこと.）この $\hat{\boldsymbol{\theta}}$ が最尤推定解に等しいことは第 7.1.7 節で示す.

7.1.5 データ多様体

EM アルゴリズムにおいては，前提として，全てのデータが計測されずに未計測のデータが存在する．計測されたデータを $\boldsymbol{y}$ として，未計測データを $\boldsymbol{x}$ とする．そして，$\{\boldsymbol{x}, \boldsymbol{y}\}$ を完全データと呼ぶ.（我々にとってなじみ深い信号源再構成問題では，$\boldsymbol{y}$ がセンサーデータ，$\boldsymbol{x}$ がボクセルデータであり，$\boldsymbol{\theta}$ はハイパーパラメータと呼ばれる．ハイパーパラメータを推定するのに EM アルゴリズムが用いられるが，このときボクセルデータ $\boldsymbol{x}$ を未知の計測データと考える.）

完全データに対する確率分布の空間

$$D = \{q(\boldsymbol{x}, \boldsymbol{y})\}$$

をデータ多様体と呼ぶ．未計測データ $\boldsymbol{x}$ が存在しない場合には，データ多様体は 1 点 $\{q(\boldsymbol{y}) : q(\boldsymbol{y}) = p(\boldsymbol{y})\}$ で構成される．しかしながら，観測されていないデータ $\boldsymbol{x}$ が存在する場合，可能性のある $\boldsymbol{x}$ に対して，$q(\boldsymbol{x}, \boldsymbol{y})$ をすべて集めたものは，S の部分多様体 D になる．ここで，$q(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{y})q(\boldsymbol{x}|\boldsymbol{y})$ として，データ多様体を

$$D = \{p(\boldsymbol{y})q(\boldsymbol{x}|\boldsymbol{y})\}$$

と表す．ここで，計測されていない量に対する確率分布が $q(\boldsymbol{x}|\boldsymbol{y})$ である．$p(\boldsymbol{y})$ は，観測値 $\boldsymbol{y}$ が定める $\boldsymbol{y}$ の分布で，$\boldsymbol{y}_0$ が観測値ならば形式的に

$$p(\boldsymbol{y}) = \delta(\boldsymbol{y} - \boldsymbol{y}_0)$$

とする．これは，経験分布とか呼ばれる確率分布で，観測データとして独立な N 個の計測値 $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N$ が求まった場合には

$$p(\boldsymbol{y}) = \sum_{i=1}^{N} \delta(\boldsymbol{y} - \boldsymbol{y}_i)$$

とする．経験分布とは，既に起きてしまった事に対する確率分布を「無理やり」定義したものである[1]．すると，データ多様体は

$$D = \{q(\boldsymbol{x}, \boldsymbol{y}) : q(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{N} \delta(\boldsymbol{y} - \boldsymbol{y}_i) q(\boldsymbol{x}|\boldsymbol{y}), \, \boldsymbol{x} \in \Omega_x\}$$

と表すことができる．ここで，Ω_x は計測されなかったデータの存在する領域である．上式で，$q(\boldsymbol{x}, \boldsymbol{y})$ は確率分布 $q(\boldsymbol{x}|\boldsymbol{y})$ の線形結合で表されるため D は混合型分布族の部分多様体である．

7.1.6 em アルゴリズム

未計測のデータ $\boldsymbol{x}$ が存在するというシナリオにおいては，モデルパラメータ $\boldsymbol{\theta}$ の推定のためにはデータの未計測部分 $\boldsymbol{x}$ をも推定しなければならない．この $\boldsymbol{x}$ と $\boldsymbol{\theta}$ の同時推定のために，データ多様体 D とモデル多様体 M を交

[1] このように定義すると，まず，$\boldsymbol{y}$ が $\boldsymbol{y}_j$ を中心とした微小区間にある確率は

$$p(\boldsymbol{y}_j - \Delta < \boldsymbol{y} < \boldsymbol{y}_j + \Delta) = \sum_{i=1}^{N} \int_{\boldsymbol{y}_j - \Delta}^{\boldsymbol{y}_j + \Delta} \delta(\boldsymbol{y} - \boldsymbol{y}_i) d\boldsymbol{y} = \int_{\boldsymbol{y}_j - \Delta}^{\boldsymbol{y}_j + \Delta} \delta(\boldsymbol{y} - \boldsymbol{y}_j) d\boldsymbol{y} = 1$$

であるので，計測されたデータを得る確率は

$$p(\boldsymbol{y}_1 - \Delta < \boldsymbol{y} < \boldsymbol{y}_1 + \Delta, \ldots, \boldsymbol{y}_N - \Delta < \boldsymbol{y} < \boldsymbol{y}_N + \Delta) = \prod_{i=1}^{N} p(\boldsymbol{y}_i - \Delta < \boldsymbol{y} < \boldsymbol{y}_i + \Delta) = \prod_{i=1}^{N} \int_{vy_i - \Delta}^{vy_i + \Delta} \delta(\boldsymbol{y} - \boldsymbol{y}_i) d\boldsymbol{y} = 1$$

となる．経験分布とは，既に起きてしまった事に対する確率が 1，それ以外のことが起こる確率 0 となる分布であり，言い換えると,「既に起きてしまったことは必ず起こり，起きなかったことは絶対に起きない」を表す確率分布である．

互に射影することで最近接点を求めるアルゴリズムが存在し，em アルゴリズムと呼ばれる.（この言葉は e-射影と m-射影を交互に折り返す事からきている．図 7.5 参照.）ただし，このノートで示すように，このアルゴリズムは EM アルゴリズムに等しく，EM アルゴリズムを通常の視点とは異なる視点で見たものに過ぎない.

em アルゴリズムの手順は以下の通りである.

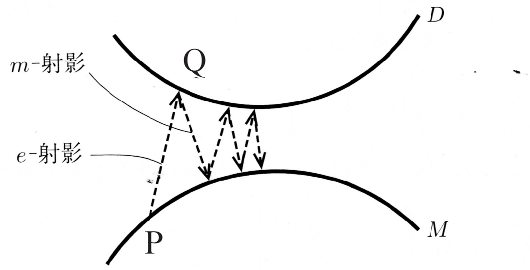

図 7.5: em アルゴリズム概念図

1. モデル多様体上に適当な初期値 $\boldsymbol{\theta}$ を選ぶ．これは，M 上に初期位置 P を $P = p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta})$ と選んだ事に対応する.

2. M 上の点 $P = p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta})$ に対するデータ多様体 D 上の e-射影 Q を求める.（ここで，D は混合型分布族であり m-平坦である.）すなわち，Q は，P から引いた e-測地線が D と直交する点，つまり，点 P からデータ多様体 D へ垂線を下した垂線の足が Q である．これは，$q(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{y})q(\boldsymbol{x})$ として，
$$\widehat{q}(\boldsymbol{x}) = \operatorname*{argmin}_{q(\boldsymbol{x})} K[p(\boldsymbol{y})q(\boldsymbol{x}) \,\|\, p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta})] = \operatorname*{argmin}_{q(\boldsymbol{x})} \int d\boldsymbol{y} d\boldsymbol{x} p(\boldsymbol{y})q(\boldsymbol{x})[\log[p(\boldsymbol{y})q(\boldsymbol{x})] - \log p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta})]$$
から $\widehat{q}(\boldsymbol{x})$ を求める事に対応する．ここで，$p(\boldsymbol{y})\widehat{q}(\boldsymbol{x})$ が D 上の点 Q を表している．このステップは e-ステップと呼ばれる.

3. 次に D 上の点 Q から，M におろした m-射影を求め，これを新しい P として P を更新する.（ここで，モデル多様体 M は指数型分布族を仮定する．したがって，e-平坦である.）すなわち，この新しい P は Q から引いた m-測地線が M と直交する点である．つまり，Q から M へ下した垂線の足が P である．これは，ステップ 2 で求まった $\widehat{q}(\boldsymbol{x})$ を $q(\boldsymbol{x})$ として，
$$\hat{\boldsymbol{\theta}} = \operatorname*{argmin}_{\boldsymbol{\theta}} K[p(\boldsymbol{y})q(\boldsymbol{x}) \| p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta})] = \operatorname*{argmin}_{\boldsymbol{\theta}} \int d\boldsymbol{y} d\boldsymbol{x} p(\boldsymbol{y})q(\boldsymbol{x})[\log[p(\boldsymbol{y})q(\boldsymbol{x})] - \log p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta})]$$
を求め，上式で求まった $\hat{\boldsymbol{\theta}}$ を新しい $\boldsymbol{\theta}$ の値としてアップデートする事に対応する．このステップは m-ステップと呼ばれる.

4. 新しい $\boldsymbol{\theta}$ の値を用いてステップ 2 に戻り，上に述べた e-射影と m-射影のステップを繰り返す.

7.1.7 em アルゴリズムと EM アルゴリズムの等価性

e ステップ（E ステップ）

まず，現在のパラメータ $\boldsymbol{\theta}$ が $(k-1)$ 回の更新の結果，$\boldsymbol{\theta}^{(k-1)}$ であったとする．このパラメータを持った確率分布が表す，モデル多様体 M の点 P は $p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})$ で表される．この P に対するデータ多様体 D 上の e-射影 Q を求める．点 Q を表す確率分布 $p(\boldsymbol{y})q(\boldsymbol{x})$ は KL ダイバージェンスを

$$K[p(\boldsymbol{y})q(\boldsymbol{x}) \,\|\, p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})] = \int d\boldsymbol{y}d\boldsymbol{x}p(\boldsymbol{y})q(\boldsymbol{x})[\log[p(\boldsymbol{y})q(\boldsymbol{x})] - \log p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})] \tag{7.2}$$

と定義して，この KL ダイバージェンスを最小とする $q(\boldsymbol{x})$ として，つまり，

$$\widehat{q}(\boldsymbol{x}) = \operatorname*{argmin}_{q(\boldsymbol{x})} K[p(\boldsymbol{y})q(\boldsymbol{x}) \,\|\, p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})]$$

から求める．これは，$q(\boldsymbol{x})$ と $p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})$ 間の KL ダイバージェンスを

$$K[q(\boldsymbol{x}), p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})] = \int d\boldsymbol{x}q(\boldsymbol{x})[\log q(\boldsymbol{x}) - \log p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})] \tag{7.3}$$

と定義して，

$$\widehat{q}(\boldsymbol{x}) = \operatorname*{argmin}_{q(\boldsymbol{x})} K[q(\boldsymbol{x}), p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})]$$

を計算するのと等価である.（なぜならば，$p(\boldsymbol{y})$ は経験分布であるので，$\boldsymbol{y}$ に関する積分は $\boldsymbol{y}$ の計測値を $\boldsymbol{y}$ に代入することに等しいだけである.）

若干長い導出（この導出は第 7.1.7 節に示す）の後，この $\widehat{q}(\boldsymbol{x})$ は

$$\widehat{q}(\boldsymbol{x}) = \frac{p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})}{p(\boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})} = p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(k-1)}) \tag{7.4}$$

と求まる．すなわち，式 (7.3) に示す KL ダイバージェンスの最小化により，その時点でのパラメータの推定値 $\boldsymbol{\theta}^{(k-1)}$ を用いた事後分布 $p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(k-1)})$ が求まる．すなわち，この e-ステップは EM アルゴリズムの E ステップに等しい．

m ステップ（M ステップ）

確率分布 $\widehat{q}(\boldsymbol{x}) = p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(k-1)})$ で表されたデータ多様体 D 上の点 Q に対応した，M 上の m-射影を求め P を更新する．この計算は E ステップで求まった $\widehat{q}(\boldsymbol{x}) = p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(k-1)})$ を用いて，

$$\hat{\boldsymbol{\theta}} = \operatorname*{argmin}_{\boldsymbol{\theta}} K[p(\boldsymbol{y})\widehat{q}(\boldsymbol{x})\|p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta})] = \operatorname*{argmin}_{\boldsymbol{\theta}} K[p(\boldsymbol{y})p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(k-1)})\|p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta})]$$

を計算するのであるが,（$\boldsymbol{y}$ に関する積分は $\boldsymbol{y}$ の計測値を $\boldsymbol{y}$ に代入することに等しいので,）結局，KL ダイバージェンスを

$$K[p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(k-1)}), p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta})] = \int d\boldsymbol{x}p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(k-1)})[\log p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(k-1)}) - \log p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta})] \tag{7.5}$$

として，これを最小とする $\boldsymbol{\theta}$ を求める事に等しい．すなわち，

$$\boldsymbol{\theta}^{(k)} = \operatorname*{argmin}_{\boldsymbol{\theta}} K[p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}^{(k-1)}), p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta})] \tag{7.6}$$

として，パラメータ $\boldsymbol{\theta}$ を $\boldsymbol{\theta}^{(k-1)} \to \boldsymbol{\theta}^{(k)}$ とアップデートする．

このとき，上式右辺の KL ダイバージェンスを以下のように変形する．まず，平均データ尤度 $\Phi(\boldsymbol{\theta})$ が

$$\Phi(\boldsymbol{\theta}) = \int d\boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)}) \log p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}) \tag{7.7}$$

で表されることに留意する．また，事後確率分布に関するエントロピー $\mathcal{H}[p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)})]$ は

$$\mathcal{H}[p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)})] = -\int d\boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)}) \log p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)}) \tag{7.8}$$

となるので，式 (7.6) に示す KL ダイバージェンスは，

$$K[p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)}), p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta})] = -\Phi(\boldsymbol{\theta}) - \mathcal{H}[p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)})] \tag{7.9}$$

と表される．つまり，この KL ダイバージェンスは，平均データ尤度とエントロピーの和に − 符号を付けたものとして表すことができる．

ここで，エントロピーの項は $\boldsymbol{\theta}$ を含まないので，$\boldsymbol{\theta}$ に関する最小化には関係しない．したがって，KL ダイバージェンス $K[p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)}), p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta})]$ を最小とするパラメータ $\boldsymbol{\theta}$ の値は，平均データ尤度 $\Phi(\boldsymbol{\theta})$ を最大とする $\boldsymbol{\theta}$ の値に等しい．すなわち，この過程は EM アルゴリズムの M ステップに等しい．

まとめ

以上をまとめると，モデル多様体 M 上の点 P：$p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})$ に対応したデータ多様体 D 上の e-射影 Q に対応する確率分布 $q(\boldsymbol{x})$ は，その時点でのハイパーパラメータ $\boldsymbol{\theta}$ の推定値 $\boldsymbol{\theta}^{(k-1)}$ を用いた事後分布 $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)})$ に等しい．すなわち，事後分布 $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)})$ を求めるこのステップは EM アルゴリズムの E ステップに等しい．

次に，データ多様体 D 上の点 Q：$p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)})$ に対応したモデル多様体 M 上の m-射影を求め，新しい P として更新する．このときの更新されたパラメータ値 $\boldsymbol{\theta}^{(k)}$ は，事後分布 $p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})$ に関する平均データ尤度を最大とする $\boldsymbol{\theta}$ に等しい．すなわち，これは EM アルゴリズムの M ステップに等しい．

このように e ステップと m ステップを交互に繰り返す em アルゴリズムは EM アルゴリズムと全く等価である．

ここで，K 回目の e-ステップ終了時点では，$q(\boldsymbol{x})$ は事後分布 $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)})$ に等しくなるため，この時点での KL ダイバージェンスは

$$\begin{aligned}
K[p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)}), p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})] &= \int d\boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)})[\log p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)}) - \log p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})] \\
&= -\int d\boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)}) \log \frac{p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})}{p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)})} = -\int d\boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)}) \log p(\boldsymbol{y}|\boldsymbol{\theta}^{(k-1)}) \\
&= -\log p(\boldsymbol{y}|\boldsymbol{\theta}^{(k-1)}) \int d\boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)}) = -\log p(\boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})
\end{aligned} \tag{7.10}$$

となる．すなわち，e-ステップ終了時点では，KL ダイバージェンスの値は，周辺尤度 $\log p(\boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})$ に − 符号をつけたものに等しくなっている．したがって，KL ダイバージェンス最小で求めた推定解はこの時点で尤度 $\log p(\boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})$ を最大とする推定解である．したがって，em(EM) アルゴリズムの更新が進む中で推定解は最尤推定解に収束していくことがわかる．

式 (7.4) の導出：

汎関数 $K[q(\boldsymbol{x}), p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})]$ を関数 $q(\boldsymbol{x})$ に関して最大とするのだが，関数 $q(\boldsymbol{x})$ はそもそも確率分布であるため制約条件，$\int_{-\infty}^{\infty} q(\boldsymbol{x})d\boldsymbol{x} = 1$ が存在する．したがって，この最適化問題は

$$\widehat{q}(\boldsymbol{x}) = \underset{q(\boldsymbol{x})}{\operatorname{argmax}}\, K[q(\boldsymbol{x}), p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})] \quad \text{subject to} \quad \int_{-\infty}^{\infty} q(\boldsymbol{x})d\boldsymbol{x} = 1 \tag{7.11}$$

となる．制約付き最適化問題であるので，ラグランジェ未定定数法を用いて無制約最適化問題に置き換えて解く．つまり，ラグランジェ未定定数を γ として，ラグランジアンを

$$\begin{aligned}\mathbb{L}[q,\gamma] &= K[q(\boldsymbol{x}), p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})] + \gamma \left[\int_{-\infty}^{\infty} q(\boldsymbol{x})d\boldsymbol{x} - 1\right] \\ &= \int_{-\infty}^{\infty} d\boldsymbol{x} q(\boldsymbol{x})[\log p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}^{(k-1)}) - \log q(\boldsymbol{x})] + \gamma \left[\int_{-\infty}^{\infty} q(\boldsymbol{x})d\boldsymbol{x} - 1\right]\end{aligned} \tag{7.12}$$

と定義する．そして，$\mathbb{L}[q,\gamma]$ を $q(\boldsymbol{x})$ と γ について微分しゼロと置くことにより最適解を求める．

まず，$q(\boldsymbol{x})$ について汎関数 $\mathbb{L}[q,\gamma]$ の微分を行い，その微係数をゼロとすれば，

$$\frac{\partial \mathbb{L}[q(\boldsymbol{x}),\gamma]}{\partial q(\boldsymbol{x})} = \log p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}^{(k-1)}) - \log q(\boldsymbol{x}) - 1 + \gamma = 0 \tag{7.13}$$

を得る．また，$\mathbb{L}[q,\gamma]$ を γ について微分し，その微係数をゼロとすれば，

$$\frac{\partial \mathbb{L}[q(\boldsymbol{x}),\gamma]}{\partial \gamma} = \int_{-\infty}^{\infty} q(\boldsymbol{x})d\boldsymbol{x} - 1 = 0 \tag{7.14}$$

を得る．式 (7.13) から，

$$\widehat{q}(\boldsymbol{x}) = e^{\gamma-1} p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}^{(k-1)}) \tag{7.15}$$

を得，また式 (7.14) から，

$$\int_{-\infty}^{\infty} q(\boldsymbol{x})d\boldsymbol{x} = e^{\gamma-1} \int_{-\infty}^{\infty} p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})d\boldsymbol{x} = e^{\gamma-1} p(\boldsymbol{y}|\boldsymbol{\theta}^{(k-1)}) = 1 \tag{7.16}$$

となるので，したがって，

$$e^{\gamma-1} = \frac{1}{p(\boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})} \tag{7.17}$$

を得る．式 (7.15) に代入すれば，結局，

$$\widehat{q}(\boldsymbol{x}) = \frac{p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})}{p(\boldsymbol{y}|\boldsymbol{\theta}^{(k-1)})} = p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)})$$

すなわち，式 (7.4) を得る．

7.2　前回の説明への補足

7.2.1　前回の説明の問題点

指数型分布族の多様体 S を考え，その内部に部分多様体 M を考える．前回の em アルゴリズムの議論で出てきたモデル多様体やデータ多様体がこれに該当する．前回の議論において，「S は指数型分布族であるので双対平坦で

あるので，M も S に含まれる部分多様体であるので，双対平坦である」と説明したがこれは間違いである．部分多様体は，それが含まれる（親）多様体のすべての性質を引き継ぐものではない．

以下，部分多様体の持つ性質について説明を行う．

7.2.2　部分多様体の定義

定義 m, n は自然数（$m < n$）とする．n 次元多様体 N の部分集合 M が N の m 次元部分多様体であるとは，M の任意の点 p の周りに，N の座標近傍 $(U; x^1, \ldots, x^n)$ が取れて，$M \cap U$ では $x^{m+1} = \cdots = x^n = 0$ となることである．このような座標近傍が見つけられる場合，M は N の部分多様体である．要するに，局所座標系を選べば $(x^1, \ldots, x^m, 0, \ldots, 0)$ と表せるような N の点の集合が，N の部分多様体である．

例 1　$N = \mathbb{R}^2$ としたとき，$y = x^2$ のグラフ上の点 M は N の部分多様体である．なぜなら，

$$\xi = x$$
$$\eta = y - x^2$$

とすれば，M の点は必ず $(\xi, 0)$ を満たす．

例 2　$N = \mathbb{R}^3$ の部分集合である半径 a の球面 M は N の部分多様体である．この場合

$$M = \{(x, y, z) \in \mathbb{R}^3 \,:\, x^2 + y^2 + z^2 = a^2\}$$

である．簡単のため第 1 象限だけで考えて，N の標準座標 (x, y, z) に対して，

$$r = \sqrt{x^2 + y^2 + z^2}$$
$$\theta = \arccos(z/r)$$
$$\phi = \arctan(y/x)$$

で決まる (θ, ϕ, r) 座標を考え，さらに $\xi = r - a$ とした座標 (θ, ϕ, ξ) を考える．この座標系で M 上の点は $(\theta, \phi, \xi) = (\theta, \phi, 0)$ と表されるので，M は N の部分多様体である．

7.2.3　部分多様体における座標系と計量

座標系：親多様体 N の点を $(U \;:\; x^1, \ldots, x^m, x^{m+1}, \ldots, x^n)$ と表す局所座標系において $U \cap M$ の各点は $(x^1, \ldots, x^m, 0, \ldots, 0)$ と表されるすれば，$U \cap M \,:\, (x^1, \ldots, x^m, 0, \ldots, 0)$ が，M の点を $\mathbb{R}^m$ の各点 $(x^1, \ldots, x^m)$ に対応付ける局所座標系になっている．

計量：親多様体 N が計量 g を持ったリーマン多様体である場合，M の計量を $\widetilde{g}$ とすると，$\widetilde{g}$ は

$$\widetilde{g}(\boldsymbol{v}, \boldsymbol{w}) = g(\boldsymbol{v}, \boldsymbol{w}) \qquad \boldsymbol{v}, \boldsymbol{w} \in T_pM \tag{7.18}$$

として自然に定義できる．ここで，$T_pM \subset T_pN$ は成り立っていて $\boldsymbol{v}, \boldsymbol{w} \in T_pN$ であることを用いている．つまり（親）多様体からの誘導計量として部分多様体の計量を自然に定義できる．（“自然に”とは「（親）多様体の計量をそのまま使ってしまえ」と言うこと．）

7.2.4 部分多様体における接続

N にアファイン接続が入っているものとする．M のベクトル場 $X, Y \in \mathfrak{X}(M)$ は N のベクトル場でもあるので，N の接続 ∇ を使って

$$\nabla_X Y$$

を計算することはできる．しかし，M の各点 $p \in M$ で，$(\nabla_X Y)_p$ は N の接ベクトルになっても，M の接ベクトルになる保証はない．つまり，一般的には部分多様体に自然に接続を誘導できない．(「(親）多様体の接続をそのまま使ってしまえ」と言うことはできない.）

これのわかりやすい例が，3 次元ユークリッド空間 N に埋め込まれた 2 次元球面 M の場合である．この 2 次元球面は 3 次元ユークリッド空間 N の部分多様体である．M の計量を式 (7.18) に示す誘導計量で定義すると，N の接続 ∇ を用いた場合，$p \in M$ に対して，$(\nabla_X \boldsymbol{Y})_p$ は，一般的には球面の接平面にはおさまらず，M の接空間からはみ出してしまう．

しかしながら，$(\nabla_X Y)_p$ が M の接ベクトルにもなっている部分多様体が存在する．このような多様体を自己平行部分多様体と呼ぶ.

定義–自己平行部分多様体： N をアファイン接続 ∇ を持つ多様体，M を N の部分多様体とする．$\forall p \in M$ と $X, Y \in \mathfrak{X}(M)$ に対して，

$$(\nabla_X Y)_p \in T_p M$$

であるとき，M は接続 ∇ に対する N の自己平行部分多様体と呼ぶ.（親多様体 N の接続が，N の部分多様体 M でも使用可能であるような M をこう呼ぶ.）

関連のある次の概念も導入する.

定義–全測地的部分多様体： N をアファイン接続 ∇ を持つ多様体，M を N の部分多様体とする．$\forall p \in M$，および，$p(0) = p$ かつ $\dot{p}(0) \in T_p M$ を満たす N の任意の $\nabla-$ 測地線 $p(t)$ が，十分小さなすべての t に対して $p(t) \in M$ となるとき，M は接続 ∇ に関する N の全測地的部分多様体であるという.

部分多様体が全測地的であるとは，$p \in M$ を通り，p で M に接する N の測地線が p の周辺でも M の中に留まっていると言うことである.

「自己平行」であることと「全測地的」であることの関係に関して，次の 2 つの定理がある.

定理 N をアファイン接続 ∇ を持つ多様体，M を N の部分多様体とする．M が自己平行であるなら，M は全測地的である.

証明： M が自己平行であるとは，N のある座標近傍 $(U; x^1, \ldots, x^m, x^{m+1}, \ldots, x^n)$ において，$M \cap U$ では $(x^1, \ldots, x^m, 0, \ldots, 0)$ なり，このとき接続係数 $\Gamma^k{}_{ij}$ が, 全ての $i, j = 1, \ldots, m$ と $m+1 \le k \le n$ に対して M 上で $\Gamma^k{}_{ij} = 0$ となることを意味する.

したがって，N の測地線方程式は

$$\ddot{x}^k(t) + \Gamma^k{}_{ij}(p)\dot{x}^i(t)\dot{x}^j(t) = 0$$

であるが，$m+1 \le k \le n$ に対して $\Gamma^k{}_{ij} = 0$ なので，

$$\ddot{x}^k(t) = 0 \qquad (k = m+1, \ldots, n)$$

である．この常微分方程式の解は積分定数を A, C として，

$$x^k(t) = At + C$$

であるが，$m+1 \leq k \leq n$ に対して，初期条件は

$$x^k(0)=0 \qquad \dot{x}^k(0)=0$$

であるので[2]，結局，$m+1 \leq k \leq n$ に対して

$$x^k(t)=0$$

を得る．これは，測地線が M 上に存在する事を意味する．

この定理の逆は成り立たないが，N の捩率がゼロである場合には，逆が成り立つ．

定理 N をアフィン接続 ∇ を持つ多様体，M を N の部分多様体とする．N の捩率がゼロ，かつ，M が全測地的であるなら，M は自己平行である．

証明：　M を局所的に $(x^1,\ldots,x^m,0,\ldots,0)$ と表す N の座標近傍 $(U;x^1,\ldots,x^m,x^{m+1},\ldots,x^n)$ を用いる．

M が全測地的であることから，$p \in M$ を通り，$\dot{p}(0) \in T_pM$ を満たす，N の測地線は全ての t に対して恒等的に

$$x^k(t)=0 \qquad (m+1 \leq k \leq n)$$

を満たす．したがって，$m+1 \leq k \leq n$ に対して $\ddot{x}^k(t)=0$ であるので，測地線方程式は

$$\Gamma^k{}_{ij}\,(p(t))\dot{x}^i(t)\dot{x}^j(t)=0 \qquad (m+1 \leq k \leq n)$$

となる．i,j に関する和の範囲を $1 \leq i,j \leq m$ に制限し，$t=0$ とすれば，任意の $p=p(0)$ において

$$\sum_{i,j=1}^{m}\Gamma^k{}_{ij}\,(p(0))\dot{x}^i(0)\dot{x}^j(0)=\sum_{i<j,i=1}^{m}\left(\Gamma^k{}_{ij}\,(p)+\Gamma^k{}_{ji}\,(p)\right)\dot{x}^i(0)\dot{x}^j(0)=0 \qquad (m+1 \leq k \leq n)$$

となる，これが任意の $\dot{x}^i(0)$（$i=1,\ldots,m$）について成り立つので，結局，各係数がゼロ，すなわち

$$\Gamma^k{}_{ij}\,(p)+\Gamma^k{}_{ji}\,(p)=0$$

が成り立つ．ここで，捩率 $T=0$ より，

$$\Gamma^k{}_{ij}\,(p)=\Gamma^k{}_{ji}\,(p)$$

であるので，上 2 つの式より，$\forall p \in M$ において

$$\Gamma^k{}_{ij}\,(p)=0 \qquad (i,j=1,\ldots,m) \quad (k=m+1,\ldots,n)$$

が成り立つ．これは，M が自己平行であることを意味している．**証明終**

N が $\nabla-$ 平坦であるとき，N の部分多様体 M が自己平行となるための条件が N のアフィン座標系を用いて次のように簡単に書ける．

定理 N をアフィン接続 ∇ に関して平坦な n 次元多様体として，そのアフィン座標系を (x^i) とする．N の m 次元部分多様体 M が自己平行となるための必要十分条件は M が N のアフィン座標系 (x^i) のアフィン部分空間に対

[2] 測地線は M 上の任意の点から引くと言う仮定なので，$x^{m+1},\ldots,x^n$ はゼロである．また，これら座標の値は M 上の任意の位置でゼロに固定なので $\dot{x}^{m+1},\ldots,\dot{x}^n$ もゼロである．

応すること，すなわち，M のある局所座標 (ξ^a) と N のアフィン座標系 (x^i) が

$$\begin{bmatrix} x^1 \\ \vdots \\ x^m \\ x^{m+1} \\ \vdots \\ x^n \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} \xi^1 \\ \vdots \\ \xi^m \end{bmatrix} + \boldsymbol{b} \tag{7.19}$$

の関係で結ばれていることである．ここで，$\boldsymbol{A}$ は $n \times m$ の行列で，$\mathrm{rank}(\boldsymbol{A}) = m$ であり，$\boldsymbol{b} \in \mathbb{R}^{m\times 1}$ である．

証明： 必要条件「M が自己平行なら，式 (7.19) に示す (ξ^a) が存在する」を示す．

まず，N は平坦である．したがって，M が自己平行なら，M も平坦である．(つまり，M に誘導された接続 $\widetilde{\nabla}$ の曲率も捩率もゼロである．）したがって，M 自身，$\widetilde{\nabla}$ アフィン座標系を持つ．それを (ξ^a) とする．M 上で (x^i) は (ξ^a) の関数である．

(ξ^a) はアフィン座標系であるので，

$$\widetilde{\nabla}_{\partial_a}\partial_b = \nabla_{\partial_a}\partial_b = \Gamma^c{}_{ab}\,\partial_c = 0$$

である.（ここで，$\nabla_{\partial_a}\partial_b$ の意味がよくわからない.）M 上で (ξ^a) は (x^i) の関数であるので，$\Gamma^c{}_{ab}\,\partial_c$ を (x^i) を使って表す．まず，良く知られた公式：

$$\Gamma^k{}_{ij} = \frac{\partial \xi^a}{\partial x^i}\frac{\partial \xi^b}{\partial x^j}\frac{\partial x^k}{\partial \xi^c}\Gamma^c{}_{ab} + \frac{\partial^2 \xi^c}{\partial x^i \partial x^j}\frac{\partial x^k}{\partial \xi^c}$$

を座標系を逆にすれば

$$\Gamma^c{}_{ab} = \frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b}\frac{\partial \xi^c}{\partial x^k}\Gamma^k{}_{ij} + \frac{\partial^2 x^k}{\partial \xi^a \partial \xi^b}\frac{\partial \xi^c}{\partial x^k} \tag{7.20}$$

を得る．したがって，

$$\Gamma^c{}_{ab}\,\partial_c = \left(\frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b}\frac{\partial \xi^c}{\partial x^k}\Gamma^k{}_{ij} + \frac{\partial^2 x^k}{\partial \xi^a \partial \xi^b}\frac{\partial \xi^c}{\partial x^k}\right)\partial_c = \left(\frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b}\Gamma^k{}_{ij} + \frac{\partial^2 x^k}{\partial \xi^a \partial \xi^b}\right)\frac{\partial \xi^c}{\partial x^k}\partial_c = 0$$

が成り立つ．ここで，(x^i) は N のアフィン座標系であるので，$\Gamma^k{}_{ij} = 0$ である．したがって，結局

$$\frac{\partial^2 x^k}{\partial \xi^a \partial \xi^b} = 0$$

が成り立つ．ここで，$k = 1, \ldots, n$，$a, b = 1, \ldots, m$ である．したがって，各 k について，

$$\frac{\partial x^k}{\partial \xi^a} = C$$

$$x^k = C_a \xi^a + B_a \qquad (a \text{ では和をとらない})$$

を得るので，$a = 1, \ldots, m$ まで加えて m で割れば，

$$x^k = \frac{1}{m}C_1\xi^1 + \cdots + \frac{1}{m}C_m\xi^m + \frac{1}{m}(B_1 + \cdots + B_m)$$

となり，式 (7.19) を得る．

次に，十分条件「式 (7.19) に示す (ξ^a) が存在すれば M は自己平行である」を示す．

式 (7.19) に示す (ξ^a) の接続 $\Gamma^c{}_{ab}$ は，式 (7.20) より，

$$\nabla_{\partial_a}\partial_b = \Gamma^c{}_{ab}\,\partial_c = \left(\frac{\partial^2 x^k}{\partial\xi^a\partial\xi^b} + \frac{\partial x^i}{\partial\xi^a}\frac{\partial x^j}{\partial\xi^b}\Gamma^k{}_{ij}\right)\partial_k$$

と書ける．(x^i) は N のアフィン座標系であるので，$\Gamma^k{}_{ij} = 0$ である．さらに，式 (7.19) に示す線形な関係から $\dfrac{\partial^2 x^k}{\partial\xi^a\partial\xi^b} = 0$ が成り立つ．したがって，

$$\nabla_{\partial_a}\partial_b = 0$$

が言える．すなわち，ξ^a はアフィン座標系であるので，部分多様体 M は平坦である．**よってすべての** $p \in M$ **で，** $(\nabla_{\partial_a}\partial_b)_p \in T_pM$ **であるので，**M **は自己平行である．**

最後の太字部分の論理展開は「なんか変」．結局ここでは，N が平坦ならば，ξ^a は平坦な座標系（アフィン座標系）となることを証明している．したがって，式 (7.19) に示す線形な関係があれば,（M は自己平行であると言うよりは）M は平坦であると言っていいのではないか? もちろん，M が平坦ならば，自己平行であることは間違いではないのであるが,「平坦性」は「自己平行性」より強い要請なので,「なんか変」と感じざるを得ない．

7.2.5　∇^e- 自己平行部分多様体

統計多様体 $(\mathcal{S}, g, \nabla^e, \nabla^m)$ の部分多様体を考える．ここで，S は指数型分布族とする．g はフィシャー計量，∇^e と ∇^m の双対接続を持つ統計多様体を仮定する．ここで，指数型分布族とは,

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \exp\left(\theta^i u_i(\boldsymbol{x}) - \psi(\boldsymbol{\theta}) + C(\boldsymbol{x})\right)$$

と表される確率分布族 $\{p(\boldsymbol{x}|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ を指数型分布族と呼ぶ．ここで，k 次元の場合，$k+1$ 個の関数 $u_1(\boldsymbol{x}), \ldots, u_k(\boldsymbol{x}), 1$ は 1 次独立に選ぶ.

このとき，次の定理が成り立つ.

定理 S の部分多様体 M に対し，M が ∇^e- 自己平行部分多様体であるための必要十分条件は M が指数型分布族であることである.

S は指数型分布族としているので，自然座標系に関して双対平坦である．したがって，M が ∇^e- 自己平行部分多様体であるための必要十分条件は M が S のアフィン座標系を用いて式 (7.19) のように表せることである.

証明：　まず，M が指数分布族であれば，M は $\nabla^{(e)}-$ 自己平行部分多様体であることを示す.

M が指数分布族であれば，$p \in M$ は自然座標系を ξ^i として,

$$\log p(\omega) = \sum_{i=1}^{k} \xi^i F_i(\omega) - \psi'(\boldsymbol{\theta})$$

と表され，また，当然 $p \in S$ であるので，S の自然座標系 θ^i を用いて,

$$\log p(\omega) = \sum_{i=1}^{n-1} \theta^i \delta_i(\omega) - \psi(\boldsymbol{\theta})$$

と表される．したがって,

$$\sum_{i=1}^{k} \xi^i F_i(\omega) = \sum_{i=1}^{n-1} \theta^i \delta_i(\omega) + (\psi(\boldsymbol{\theta}) - \psi'(\boldsymbol{\theta}))$$

が成り立つ．ここで，$F_i(\omega)$ は $\delta_i(\omega)$（$i=1,\ldots,n$）の線形結合で表すことができる．したがって，自然座標系 θ^i と ξ^j は

$$\begin{bmatrix} \theta^1 \\ \vdots \\ \theta^n \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} \xi^1 \\ \vdots \\ \xi^k \end{bmatrix} + \boldsymbol{b}$$

と表すことができる．$\boldsymbol{A}$ は $n\times k$ の係数行列であり，$\boldsymbol{b}$ は定数ベクトルである．すなわち，これは式 (7.19) であり，自己平行部分多様体であることが示される．ここで，$\boldsymbol{\theta}$ は $\nabla^{(e)}-$ アフィン座標系であるので，M は $\nabla^{(e)}-$ 自己平行部分多様体である．

つぎに，逆方向．M が $\nabla^{(e)}-$ 自己平行部分多様体であれば，M が指数型分布族であることを示す．

(省略）上に述べた証明を逆にたどる．面倒なので省略.

7.2.6 $\nabla^{(e)}-$ 自己平行部分多様体の双対平坦性

M を S の $\nabla^{(e)}-$ 自己平行部分多様体とすると，M 上には（自然に）計量 $\widetilde{g}$ と接続 $\widetilde{\nabla}^{(e)}$ が誘導される．

すなわち，計量は標準的な誘導計量

$$\widetilde{g}_p(\boldsymbol{v},\boldsymbol{w}) = g_p(\boldsymbol{v},\boldsymbol{w}) \qquad (\boldsymbol{v},\,\boldsymbol{w}\in T_pM)$$

を用いる.（$T_pM\subset T_pN$ であるので，$\boldsymbol{v},\,\boldsymbol{w}\in T_pN$ でもある.）

接続は，M の N から誘導された接続 $\widetilde{\nabla}$ を

$$\widetilde{g}(\widetilde{\nabla}_X\boldsymbol{Y},\boldsymbol{Z}) = g(\nabla_X\boldsymbol{Y},\boldsymbol{Z}) \qquad (\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}\in\mathfrak{X}(M))$$

から決める.

M 上に計量 $\widetilde{g}$ と接続 $\widetilde{\nabla}^{(e)}$ が誘導されると,

$$\boldsymbol{X}\widetilde{g}(\boldsymbol{Y},\boldsymbol{Z}) = \widetilde{g}(\widetilde{\nabla}^{(e)}_X\boldsymbol{Y},\boldsymbol{Z}) + \widetilde{g}(\boldsymbol{Y},\widetilde{\nabla}^*_X\boldsymbol{Z})$$

から，M の双対接続 $\widetilde{\nabla}^*$ を導入する．すると，$(\widetilde{g},\widetilde{\nabla}^{(e)},\widetilde{\nabla}^*)$ は M の双対構造となる．M が $\widetilde{\nabla}^{(e)}-$ 自己平行であったから，M の $\widetilde{\nabla}^{(e)}-$ 曲率はゼロであるので，M の $\widetilde{\nabla}^*-$ 曲率もゼロである．ここで，S の $\nabla^{(m)}-$ 捩率はゼロである．$\widetilde{\nabla}^*_X\boldsymbol{Z}$ は $\nabla^{(m)}_X\boldsymbol{Z}$ を M の接空間に g で投影したものであるので M の $\widetilde{\nabla}^*-$ 捩率もゼロとなる．結局 M は誘導された双対構造 $(\widetilde{g},\widetilde{\nabla}^{(e)},\widetilde{\nabla}^*)$ に関して双対平坦な多様体となる.

このとき，正準パラメータ $\boldsymbol{\theta}$ は M の $\widetilde{\nabla}^{(e)}-$ アフィン座標系となっている．それでは，$\widetilde{\nabla}^*-$ アフィン座標系はどう求まるのであろうか.

7.2.7 $\nabla^{(e)}-$ 自己平行部分多様体の $\widetilde{\nabla}^*-$ アフィン座標系

S の $\nabla^{(e)}-$ 自己平行部分多様体は指数型分布族

$$M = \{p(\omega)\in S\,:\,\log p(\omega) = C(\omega) + \sum_{i=1}^{k}\theta^i F_i(\omega) - \psi(\boldsymbol{\theta})\}$$

である．ここで，正準パラメータ θ^i は M の $\widetilde{\nabla}^{(e)}-$ アフィン座標系となっている．これに対し，

$$\eta_i = E[F_i(\omega)] = \sum_{\omega\in\Omega} p(\omega)F_i(\omega)$$

とすれば，$\{(\theta^i), (\eta_i)\}$ は双対構造 $(\widetilde{g}, \widetilde{\nabla}^{(e)}, \widetilde{\nabla}^*)$ に対する双対アフィン座標系を成す．

証明：写像 $\boldsymbol{\theta} \mapsto \boldsymbol{\eta}$ のヤコビ行列は，

$$\frac{\partial\eta_i}{\partial\theta^j} = \sum_{\omega\in\Omega}(\partial_j p(\omega))F_i(\omega) = \sum_{\omega\in\Omega}(\partial_j p(\omega))\,(\partial_i \log p(\omega) + \partial_i\psi(\boldsymbol{\theta})) = \sum_{\omega\in\Omega}(\partial_j p(\omega))\,(\partial_i \log p(\omega)) = \widetilde{g}_{ij}$$

となって，ヤコビ行列は計量に等しい．計量は正定値なので，ヤコビ行列も正定値であるので，η_i は θ^i から導入可能な座標系である．

さらに，

$$\widetilde{g}_{ij} = \widetilde{g}(\partial_i, \partial_j) = \widetilde{g}(\partial_i, \frac{\partial\eta_k}{\partial\theta^j}\partial^k) = \widetilde{g}(\partial_i, \partial^k)\frac{\partial\eta_k}{\partial\theta^j} = \widetilde{g}(\partial_i, \partial^k)\widetilde{g}_{kj}$$

であるので，

$$\widetilde{g}(\partial_i, \partial^k) = \delta_i^k$$

が導かれる．つまり，η_i は θ^i の双対座標系である．

さらに，

$$0 = \frac{\partial}{\partial\eta_\ell} tg(\partial_i, \partial^k) = \widetilde{g}(\widetilde{\nabla}^{(e)}_{\partial^\ell}\partial_i, \partial^k) + \widetilde{g}(\partial_i, \widetilde{\nabla}^*_{\partial^\ell}\partial^k) = \widetilde{g}(\partial_i, \widetilde{\nabla}^*_{\partial^\ell}\partial^k)$$

が成り立つ．　**（ただし，$\widetilde{\nabla}^{(e)}_{\partial^\ell}\partial_i = 0$ の理由不明．）**上式が，任意の i で成り立つので，

$$\widetilde{\nabla}^*_{\partial^\ell}\partial^k (= {\Gamma^\alpha}_{\ell k}\,\frac{\partial}{\partial\eta_\alpha}) = 0$$

を得る．したがって，$\widetilde{\nabla}^*_{\partial^\ell}\partial^k = 0$ は全ての α で接続係数 ${\Gamma^\alpha}_{\ell k}$ がゼロとなることを意味する．したがって，η^i は接続 $\widetilde{\nabla}^*$ に対するアフィン座標系である．**証明終**

7.2.8　∇^m- 自己平行部分多様体

混合型分布族：確率分布が

$$p(\omega) = \sum_{i=1}^n \eta_i p_i(\omega) + \left(1 - \sum_{i=1}^n \eta_i\right) p_0(\omega)$$

で表される確率分布族 $M = \{p(\omega|\eta) : \eta \in H\}$ を混合型分布族という．ここで，$p_0, p_1, \ldots, p_n$ は 1 次独立であるとする．

次の定理が成り立つ．

定理　S の部分多様体 M に対して，M が $\nabla^{(m)}-$ 自己平行部分多様体であるための必要十分条件は，M が混合型分布族であることである．

証明： 指数型分布族の場合とほとんど同じようにして証明できる．M が混合型分布族であれば，M の座標系を ζ^i として，$p \in M$ は

$$p(\omega) = \sum_{i=1}^{k} \zeta_i \widetilde{p}_i(\omega) + \left(1 - \sum_{i=1}^{k} \zeta_i\right) \widetilde{p}_0(\omega)$$

と表され，また，当然ながら $p \in S$ であるので，S の座標系 η^i を用いて，

$$p(\omega) = \sum_{i=1}^{n} \eta_i p_i(\omega) + \left(1 - \sum_{i=1}^{k} \eta_i\right) p_0(\omega)$$

と表される．したがって，

$$\sum_{i=1}^{k} \zeta_i \widetilde{p}_i(\omega) = \sum_{i=1}^{n} \eta^i p_i(\omega) + C$$

が成り立つ．ここで，C は，座標に依存しない定数項を表した．$\widetilde{p}_i(\omega)$ は $p_i(\omega)$ の線形結合で表されるため，結局，座標系 η_i と ζ_i の関係は

$$\begin{bmatrix} \eta^1 \\ \vdots \\ \eta^n \end{bmatrix} = \boldsymbol{B} \begin{bmatrix} \zeta^1 \\ \vdots \\ \zeta^k \end{bmatrix} + \boldsymbol{c}$$

の形で表すことができる．$\boldsymbol{B}$ はある $n \times k$ の行列であり,，$\boldsymbol{c}$ は定数ベクトルである．したがって，M は S の $\nabla^{(m)}-$ 自己平行部分多様体である．

反対に，M が $\nabla^{(m)}-$ 自己平行部分多様体であれば，座標系 η_i と ζ_j に上式の関係が成り立つので，M は混合型分布族である．詳細は省略．

7.2.9 Appendix：第 7.2.8 節の定理ー文献 [2] の証明

第 7.2.8 節において次の定理：

定理 S の部分多様体 M に対し，M が ∇^e- 自己平行部分多様体であるための必要十分条件は M が指数型分布族であることである．

を証明した．

この証明は，ほぼ文献 [1] によったものであるが，[2] では以下のように記載されている．

[2] の証明——S は指数型分布族と仮定しているので，S を構成する確率分布 $p(\omega)$ は

$$p(\omega) = \exp\left[C(\omega) + \sum_{i=1}^{n} \theta^i F_i(\omega) - \psi(\boldsymbol{\theta})\right]$$

ここで，M を S の部分多様体とする．

まず，M が指数分布族と仮定する．このとき，同じ元は M の元として，この場合の自然座標 $\boldsymbol{\zeta}$ を用いて，$k \leq n$ として

$$q(\omega) = p(\omega) = \exp\left[D(\omega) + \sum_{a=1}^{k} \zeta^a F_a(\omega) - \psi'(\boldsymbol{\zeta})\right]$$

と表せる．したがって，

$$C(\omega) + \sum_{i=1}^{n} \theta^i F_i(\omega) - \psi(\boldsymbol{\theta}) = D(\omega) + \sum_{a=1}^{k} \zeta^a F_a(\omega) - \psi'(\boldsymbol{\zeta})$$

が成り立つ．両辺を ζ^b で微分すれば，

$$\sum_{i=1}^{n} \frac{\partial \theta^i}{\partial \zeta^b} F_i(\omega) - \sum_{i=1}^{n} \frac{\partial \psi(\boldsymbol{\theta})}{\partial \theta^i} \frac{\partial \theta^i}{\partial \zeta^b} = F_b(\omega) - \frac{\partial \psi'(\boldsymbol{\zeta})}{\partial \zeta^b}$$

すなわち，

$$\sum_{i=1}^{n} \left(F_i(\omega) - \frac{\partial \psi(\boldsymbol{\theta})}{\partial \theta^i} \right) \frac{\partial \theta^i}{\partial \zeta^b} = F_b(\omega) - \frac{\partial \psi'(\boldsymbol{\zeta})}{\partial \zeta^b}$$

ここで，**関数 $F_1(\omega), \ldots, F_n, 1$ は 1 次独立である．したがって，任意の i, b に対して，$\dfrac{\partial \theta^i}{\partial \zeta^b}$ は定数関数となる．どうして??**

したがって，$\theta^1, \ldots, \theta^n$ は $\zeta_1, \ldots, \zeta_k$ の 1 次式で表される．したがって，M は S の $\nabla^{(e)}-$ 自己平行多様体である．

7.3　感想

EM アルゴリズムをモデル多様体とデータ多様体の最近接点を推定するアルゴリズムとして解釈すること，すなわち EM アルゴリズムの幾何学的新解釈を提案している．

ベイズ信号処理の第 4.6 節において，自由エネルギーと呼ばれる汎関数を最大化することで，EM アルゴリズムが導出できることを示している．このノートでの EM アルゴリズムの議論は，この自由エネルギーを用いた議論を小変更しただけのものであり，したがって，議論の筋道はそんなに目新しいものではない．

大きな違いは,（1）KL ダイバージェンスの意味の明確化と（2）最適化する汎関数の意味付けにある．（2）に関して，この情報幾何を用いた議論では，汎関数はモデル多様体とデータ多様体間の距離を表現した KL ダイバージェンスであるとの “かなり明瞭な” 意味がある．一方，ベイズ信号処理のの第 4.6 節の議論では，汎関数は「自由エネルギー」と呼ばれる，なにやら得体のしれないものである．いかにも深遠な統計物理学の世界から借用してきた概念であると思わせて，質問を封じてはいるが，あらためて「それなんですか？」と聞かれたら困ってしまうものである．

したがって，この情報幾何を用いた EM アルゴリズムの幾何学的解釈は，EM アルゴリズムの一般的説明としては悪くない説明である．

参考文献

以下の文献を参考にした.

1. 藤原章夫, “情報幾何学の基礎” 牧野書店，2015.

2. 藤岡敦, “入門情報幾何” 共立出版，2021.

3. 甘利俊一, “情報幾何学の新展開” サイエンス社，2014.

4. 小林昭七, “曲線と曲面の微分幾何” 裳華房，1977

5. 榎本一之, “多様体への道” 近代科学社，2016.

6. デイヴィッド C. ケイ，“テンソル解析” プレアデス出版　2018.

7. 甘利さん講義ビデオ，https://www.youtube.com/watch?v=kq2Y6cXl2ik

8. 甘利俊一, “情報幾何とその応用” システム・制御・情報, Vol.48, No.6, pp.227-235, 2004.

9. 赤穂昭太郎, “ EM アルゴリズムの幾何学”, 情報処理, Vol.37, No. 1, 1996